Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

DIPLOMOVÁ PRÁCE

Szilárd Kálosi

Penalizační metody ve stochastické optimalizaci

Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: RNDr. Martin Branda, Ph.D.
Studijní program: Matematika
Studijní obor: Pravděpodobnost, matematická statistika a ekonometrie

Praha 2013

I hereby express my deepest gratitude to my supervisor RNDr. Martin Branda, Ph.D. for his careful guidance, numerous comments and help throughout the elaboration, which remarkably contributed to the improvement of the thesis.
For all the support, my sincerest thanks belongs to my girlfriend and family, as well.

Prohlašuji, že jsem tuto diplomovou práci vypracoval samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova v Praze má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V Praze dne 1. srpna 2013 ....

Název práce: Penalizační metody ve stochastické optimalizaci

Autor: Szilárd Kálosi

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: RNDr. Martin Branda, Ph.D.

Abstrakt: Předložená práce se zabývá penalizační metodou ve stochastické optimalizaci. Hlavním cílem práce je studium penalizačních metod v deterministické optimalizaci, zejména exaktních penalizačních metod, za účelem rozšíření penalizačních metod ve stochastické optimalizaci. Za tímto účelem ukážeme ekvivalenci výchozího deterministického nelineárního a odpovídajícího penalizačního problému používajícího libovolnou vektorovou normu jako penalizační funkci, a to pro konvexní a invexní funkce vyskytující se v problémech. Získané věty jsou následně aplikovány na problémech s mnohonásobným pravděpodobnostním omezením s konečně diskrétním pravděpodobnostním rozdělením k dokázání asymptotické ekvivalence stochastického a odpovídajícího penalizačního problému. Praktické použití nově získaných metod je demonstrováno v numerické studii, ve které je rovněž poskytnuto srovnání s ostatnými přístupy.

Klíčová slova: asymptotické ekvivalence, konvexní funkce, invexní funkce, mnohonásobné pravděpodobnostní omezení, penalizační metody

Title: Penalty methods in stochastic optimization

Author: Szilárd Kálosi

Department: Department of Probability and Mathematical Statistics

Supervisor of the master thesis: RNDr. Martin Branda, Ph.D.

Abstract: The submitted thesis studies penalty function methods for stochastic programming problems. The main objective of the paper is to examine penalty function methods for deterministic nonlinear programming, in particular exact penalty function methods, in order to enhance penalty function methods for stochastic programming. For this purpose, the equivalence of the original deterministic nonlinear and the corresponding penalty function problem using arbitrary vector norm as the penalty function is shown for convex and invex functions occurring in the problems, respectively. The obtained theorems are consequently applied to multiple chance constrained problems under finite discrete probability distribution to show the asymptotic equivalence of the probabilistic and the corresponding penalty function problems. The practical use of the newly obtained methods is demonstrated on a numerical study, in which a comparison with other approaches is provided as well.

Keywords: asymptotic equivalence, convex functions, invex functions, multiple chance constraints, penalty function methods

Contents

Introduction

Mathematical programming deals with the problem of optimising a certain criterion expressed as an objective function over a set of feasible decisions. The set of feasible decisions is usually described by equality and inequality constraints, i.e. functions satisfying certain equality and inequality conditions, respectively. If such constraints exist, the problem is called constraint programming, otherwise the problem is called unconstrained programming.

If all the functions are nonlinear, the problem is called nonlinear programming. For solving such problems, various algorithms have been proposed, e.g. feasible direction algorithms [3] for constrained nonlinear programming and line search algorithms [3] for unconstrained nonlinear programming.

Penalty function methods are a special type of such algorithms. They convert a constrained problem into a sequence of unconstrained problems or into a single unconstrained one. The constraints of the original problem are incorporated into the objective function of the unconstrained problem through a suitably chosen penalty function, which penalises any violations of the constraints. The extent of the penalisation is controlled by a nonnegative penalty coefficient. By making the penalty coefficient larger, the violations of the constraints are naturally penalised more strictly. It can be shown that under mild assumptions, the penalty function methods are able to generate optimal solutions to the original problems by increasing the penalty coefficient.

For exterior penalty function methods, the penalty parameter must usually be infinitely large, in a limiting sense, in order to obtain an optimal solution to the original problem. This can cause computational difficulties and ill-conditioned Hessian matrix, for further details see [3] or [26]. Another weakness of the exterior penalty functions is the assumption of finding the global minimisers to the unconstrained penalty subproblems. In practise, we do not obtain global minimisers as the algorithms usually generate sequences of locally optimal solutions. Hence, exact penalty methods which attempt to solve the original problem by finding a solution of a single unconstrained problem have been employed.

In many practical cases, one must cope with the consequences of realisations of the randomness occurring during the decision making process and hedge supremely against them so as to secure the best possible outcome. Ignoring this uncertainty may lead to unsatisfactory or simply wrong decisions. Stochastic programming problems differ from deterministic problems, in which all the coefficients are exactly known, in incorporating this uncertainty into the model. Natural demand for solving such problems occurs in many different fields of science and engineering, from finance and economics to agriculture, medicine and logistics, cf. [33] or [34].

Two predominant approaches has been employed to solve the treatment of the random part. One of them is based on the notion of optimising the total lost or cost on average and called expected violation (or shortfall) penalty model. The second approach is derived from reliability requirements on the optimal decision and called probabilistic (or chance constrained) programs.

Numerical solving stochastic programming problems, in particular probabilistic programs, is decidedly more difficult than numerical solving deterministic programs. The feasible region of a probabilistic program is generally not convex, and it is not easy to check the feasibility of a point as it leads to computation of multivariate integrals. Therefore, finding reliable numerical solution to probabilistic programs has been of greatest interest. Fortunately, there exist several methods for numerically solving chance constrained problems under discrete or continuous probability distributions, cf. [28]. In case of continuous or discrete probability distribution with many realisation, approximative techniques can be used to solve the problems numerically, cf. [7].

It was previously noted that by solving problems with penalised random constraints, one can obtain highly reliable solutions to the original probabilistic program, cf. [8]. This type of penalisation was successfully applied in water management [14], insurance [15] and finance [6, 7]. These observations led to study the behaviour of penalty function methods applied on chance constrained problems with respect to their possible equivalence. Rigorous proofs of the asymptotic equivalences of chance constrained problems and penalty function problems under various assumptions were submitted in [15, 9, 7, 8].

The aim of the thesis is to research penalty function methods for deterministic nonlinear programming problems in order to enhance the applicability of these methods, specifically of exact penalty function methods, for probabilistic programming problems. The thesis also includes a numerical study on which the newly constructed theorems are compared with already available approaches to solve probabilistic programming problems.

In order to fulfil the aforementioned aims, the thesis is organised as follows:

In chapter 1, penalty function methods for deterministic nonlinear programming are delineated. In section 1.1, the original problem is formulated, and the definition of suitable penalty functions is proposed. Section 1.2 is devoted to exterior penalty function methods. The main attention is focused on section 1.3 in which three theorems concerning exact penalty function methods are put forward. The equivalence of the original and the penalty function problem using arbitrary vector norm as penalty function is shown for convex and invex functions, respectively.

In chapter 2, the reader is acquainted with the basic notions of one-stage stochastic programming problems. The two prevalent approaches to treat the uncertainty are briefly outlined in sections 2.1 and 2.2.

In chapter 3, penalty function methods for stochastic programming problems are described. The chapter is divided into two sections. Section 3.1 discusses the asymptotic equivalence of the multiple chance constrained and the corresponding penalty function problems under continuous probability distributions. Section 3.2 studies the same asymptotic equivalence under finite discrete probability distributions. In this section, new penalty function methods for solving multiple chance constrained problems, based on theorems proposed in section 1.3, are put forward as well.

In chapter 4, a numerical study demonstrating the capability of the newly proposed methods to generate highly reliable solutions with other methods is provided. The usefulness of the newly proposed penalty function methods for stochastic programming problems is illustrated on a VaR-constrained portfolio selection problem [2].

In Appendix A, the less common terms and relations used throughout the thesis are gathered together and defined or derived for convenience.

Chapter 1

Penalty function methods for nonlinear programming

In this chapter, we introduce the basic concept of penalty function method and present the main theorem concerning its convergence to the original problem. Furthermore, we introduce exact penalty function methods for convex and invex functions, which do not require to solve an infinite sequence of unconstrained nonlinear programming problems.

1.1 Notations and notions

First, we formulate the constrained nonlinear problem and propose a proper definition of suitable penalty functions for the original problem.

1.1.1 The original problem

The formulation of the original problem follows the formulation of the primal problem in [3, p. 476]. Let $\mathcal{X}$ be a nonempty set in $\mathbb{R}^n$. Consider the following minimisation problem

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & \boldsymbol{g}(\boldsymbol{x}) \leq \mathbf{0} \\ & \boldsymbol{h}(\boldsymbol{x}) = \mathbf{0} \\ & \boldsymbol{x} \in \mathcal{X}, \end{aligned} \tag{OP}$$

where we put $\boldsymbol{g}(\boldsymbol{x}) = (g_1(\boldsymbol{x}), \dots, g_m(\boldsymbol{x}))^T$, $\boldsymbol{h}(\boldsymbol{x}) = (h_1(\boldsymbol{x}), \dots, h_l(\boldsymbol{x}))^T$ and $\mathbf{0} = (0, \dots, 0)^T$ for convenience. Hereinafter, the problem (OP) will be referred to as the original problem. The functions $f(\boldsymbol{x})$, $g_1(\boldsymbol{x}), \dots, g_m(\boldsymbol{x})$, $h_1(\boldsymbol{x}), \dots, h_l(\boldsymbol{x})$ are defined on $\mathbb{R}^n$ and continuous in their arguments, $f(\boldsymbol{x}) : \mathbb{R}^n \to \mathbb{R}$, $\boldsymbol{g}(\boldsymbol{x}) : \mathbb{R}^n \to \mathbb{R}^m$, and $\boldsymbol{h}(\boldsymbol{x}) : \mathbb{R}^n \to \mathbb{R}^l$. Denote

$$S = \{\boldsymbol{x} \in \mathcal{X} \colon \boldsymbol{g}(\boldsymbol{x}) \leq \mathbf{0}, \boldsymbol{h}(\boldsymbol{x}) = \mathbf{0}\}$$

the feasible region of the original problem (OP). The set $\mathcal{X}$ might typically represent constraints which could be handled directly, e.g lower and upper bounds on the variables.

For further needs, let us define the Lagrangian function for the original problem (OP), which is defined as

$$\mathcal{L}(\boldsymbol{x}; \boldsymbol{u}, \boldsymbol{v}) = \begin{cases} f(\boldsymbol{x}) + \sum_{i=1}^{m} u_i g_i(\boldsymbol{x}) + \sum_{i=1}^{l} v_i h_i(\boldsymbol{x}), & \text{if } \boldsymbol{x} \in \mathcal{X}, \\ \infty, & \text{if } \boldsymbol{x} \notin \mathcal{X} \end{cases}$$

where $\boldsymbol{u} = (u_1, \ldots, u_m)^T$, $u_i \geq 0$, $i = 1, \ldots, m$, and $\boldsymbol{v} = (v_1, \ldots, v_l)^T$, $v_i \in \mathbb{R}$, $i = 1, \ldots, l$, are called Lagrangian multipliers associated with the inequality $g_i(\boldsymbol{x})$, $i = 1, \ldots, m$, and the equality constraints $h_i(\boldsymbol{x})$, $i = 1, \ldots, l$, respectively.

1.1.2 Penalty functions

A suitable penalty function must *vanish* for feasible points and return *positive* and *negative* values for infeasible points of minimisation and maximisation problems, respectively, cf. [3]. We propose the following definition of penalty functions based on the requirements stated above.

Definiton 1.1. A function $p(\boldsymbol{x}) : \mathbb{R}^n \to \mathbb{R}$ is called a *penalty function* for the original problem (OP), if $p(\boldsymbol{x})$ satisfies

(i) $p(\boldsymbol{x}) = 0$ for $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}$ and $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}$,

(ii) $p(\boldsymbol{x}) > 0$ for $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{g}(\boldsymbol{x}) > \boldsymbol{0}$ or $\boldsymbol{h}(\boldsymbol{x}) \neq \boldsymbol{0}$.

Having stated the definition of the penalty function $p(\boldsymbol{x})$, we present the basic penalty function method form of the original problem (OP). Denote μ the penalty coefficient. Naturally, it must hold that $\mu \geq 0$. Thus, the penalty function method form of the original problem (OP) using the penalty function $p(\boldsymbol{x})$ is formulated as follows

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) + \mu p(\boldsymbol{x}) \\ \text{subject to} \quad & \boldsymbol{x} \in \mathcal{X}. \end{aligned} \tag{PP}$$

Penalty functions $p(\boldsymbol{x})$ are typically defined by

$$p(\boldsymbol{x}) = \sum_{i=1}^{m} \phi(g_i(\boldsymbol{x})) + \sum_{j=1}^{l} \psi(h_j(\boldsymbol{x})), \tag{1.1}$$

where $\phi(y)$ and $\psi(y)$ are functions such that

$$\begin{aligned} \phi(y) = 0 \quad \text{for } y \leq 0, \qquad & \text{and} \qquad \phi(y) > 0 \quad \text{for } y > 0, \\ \psi(y) = 0 \quad \text{for } y = 0, \qquad & \text{and} \qquad \psi(y) > 0 \quad \text{for } y \neq 0, \end{aligned} \tag{1.2}$$

although more general function satisfying the definition can be conceptually used. Typically, $\phi(y)$ and $\psi(y)$ are of the forms

$$\phi_q(y) = \max\{0, y\}^q, \quad \text{and} \quad \psi_q(y) = |y|^q, \tag{1.3}$$

where q is positive, cf. [3]. Let us define the L_q norm for a given vector $\boldsymbol{x} \in \mathbb{R}^n$ and a given $q \geq 1$ by

$$\|\boldsymbol{x}\|_q = (|x_1|^q + |x_2|^q + \cdots + |x_n|^q)^{\frac{1}{q}},$$

and the maximum norm L_∞ by

$$\|\boldsymbol{x}\|_\infty = \max\{|x_1|, |x_2|, \ldots, |x_n|\}. \tag{1.4}$$

By substituting (1.3) into (1.1) and noting that $\sum_{j=1}^{l} |h_j(\boldsymbol{x})|^q = \|\boldsymbol{h}(\boldsymbol{x})\|_q^q$ and $\sum_{i=1}^{m} [\max\{0, g_i(\boldsymbol{x})\}]^q = \|\max\{\mathbf{0}, \boldsymbol{g}(\boldsymbol{x})\}\|_q^q$, we have

$$p_q(\boldsymbol{x}) = \sum_{i=1}^{m} [\max\{0, g_i(\boldsymbol{x})\}]^q + \sum_{j=1}^{l} |h_j(\boldsymbol{x})|^q = \|\max\{\mathbf{0}, \boldsymbol{g}(\boldsymbol{x})\}\|_q^q + \|\boldsymbol{h}(\boldsymbol{x})\|_q^q. \tag{1.5}$$

The function $p_q(\boldsymbol{x})$ are generally called the L_q penalty functions, derived from their association with the L_q norm.

1.2 Exterior penalty function methods

Define the following auxiliary function

$$\theta(\mu) = \inf\{f(\boldsymbol{x}) + \mu p(\boldsymbol{x}) : \boldsymbol{x} \in \mathcal{X}\}$$

of the penalty coefficient μ. Then the penalty function method (PP) can be alternatively reformulated as follows

$$\begin{array}{rl} \sup & \theta(\mu) \\ \text{subject to} & \mu \geq 0. \end{array} \tag{APP}$$

It can be shown that under certain assumptions, the two problems (OP) and (APP) are equivalent. To show the equivalency, the following lemma is necessary.

Lemma 1.1. *Suppose that $f(\boldsymbol{x})$, $g_1(\boldsymbol{x}), \ldots, g_m(\boldsymbol{x})$, $h_1(\boldsymbol{x}), \ldots, h_l(\boldsymbol{x})$ are continuous functions on $\mathbb{R}^n$, and let $\mathcal{X}$ be a nonempty set in $\mathbb{R}^n$. Let $p(\boldsymbol{x})$ be a continuous function defined by Definition 1.1, and suppose that for each μ, there exists an $\boldsymbol{x}_\mu \in \mathcal{X}$ such that $\theta(\mu) = f(\boldsymbol{x}_\mu) + \mu p(\boldsymbol{x}_\mu)$. Then, the following statements hold true:*

(i) $\inf\{f(\boldsymbol{x}) : \boldsymbol{x} \in \mathcal{X}, \boldsymbol{g}(\boldsymbol{x}) \leq \mathbf{0}, \boldsymbol{h}(\boldsymbol{x}) = \mathbf{0}\} \geq \sup_{\mu \geq 0} \theta(\mu)$,

(ii) $f(\boldsymbol{x}_\mu)$ is a nondecreasing function of $\mu \geq 0$, $\theta(\mu)$ is a nondecreasing function of μ, and $p(\boldsymbol{x}_\mu)$ is a nonincreasing function of μ.

Proof. See [3, lemma 9.2.1]. □

Having stated the previous lemma, we can present the main theorem of section 1.2.

Theorem 1.1. *Consider the original problem* (OP)*. Suppose that the problem has a feasible solution, and let $p(\boldsymbol{x})$ be a continuous function satisfying Definition 1.1. Furthermore, suppose that for each $\mu \geq 0$ there is a solution $\boldsymbol{x}_\mu \in \mathcal{X}$ to the penalty function problem* (PP)*, and that $\{\boldsymbol{x}_\mu\}$ is contained in a compact subset of $\mathcal{X}$. Then*

$$\inf\{f(\boldsymbol{x}) : \boldsymbol{g}(\boldsymbol{x}) \leq \mathbf{0}, \boldsymbol{h}(\boldsymbol{x}) = \mathbf{0}, \boldsymbol{x} \in \mathcal{X}\} = \sup_{\mu \geq 0} \theta(\mu) = \lim_{\mu \to \infty} \theta(\mu).$$

Moreover, the limit $\bar{\boldsymbol{x}}$ of any convergent subsequence of $\{\boldsymbol{x}_\mu\}$ is an optimal solution to the original problem (OP)*, and $\mu p(\boldsymbol{x}_\mu) \to 0$ as $\mu \to \infty$.*

Proof. See [3, Theorem 9.2.2]. □

Note the assumption requiring that the sequence $\{\boldsymbol{x}_\mu\}$ of the solutions to the penalty function problems (PP) does not restrict the applicability of the theorem in many practical cases seeing that the variables are usually bounded.

It is evident from Theorem 1.1 that if $p(\boldsymbol{x}_\mu) = 0$ for some $\mu \geq 0$, then $\boldsymbol{x}_\mu$ is an optimal solution to the original problem (OP). In other words, if the penalty term $p(\boldsymbol{x}_\mu)$ vanishes for a sufficiently large penalty parameter μ, an optimal solution to the original problem (OP) is obtained. From Theorem 1.1, it also follows that the optimal solutions $\boldsymbol{x}_\mu$ to the penalty function problems (PP) can be made arbitrarily close to an optimal solution to the original problem (OP) by choosing μ large enough. Similarly, $f(\boldsymbol{x}_\mu) + \mu p(\boldsymbol{x}_\mu)$ can be made arbitrarily close to the optimal objective value of the original problem (OP). The points $\{\boldsymbol{x}_\mu\}$ are generally infeasible and approach an optimal solution from outside the set of feasible solutions by increasing the penalty coefficient μ. Hence, this technique is also denominated as *exterior* penalty function method, cf. [3].

1.2.1 Estimation of the KKT Lagrange multipliers at optimality

Under the assumptions of theorem 1.1, the KKT Lagrange multipliers associated with the constraints at optimality can be recovered by using the solutions $\{\boldsymbol{x}_\mu\}$ to the penalty subproblems (PP). The following part is partially based on [3, pp. 479-481] and exercise 9.12 in [3, p. 523].

Suppose that the penalty function is given by (1.1),

$$\mathcal{X} = \{\boldsymbol{x} \in \mathbb{R}^n \colon g_i(\boldsymbol{x}) \leq 0,\ i = m+1, \ldots, M,\ h_i(\boldsymbol{x}) = 0,\ i = l+1, \ldots, L\}, \tag{1.6}$$

and $g_{m+1}(\boldsymbol{x}), \ldots, g_M(\boldsymbol{x}),\ h_{l+1}(\boldsymbol{x}), \ldots, h_L(\boldsymbol{x})$ are differentiable. Note that the penalty function $p(\boldsymbol{x})$ might not be continuously differentiable, e.g. functions such as $\max\{0, g_i(\boldsymbol{x})\}$ are generally not differentiable at points $\boldsymbol{x}$ where $g_i(\boldsymbol{x}) = 0$. Nevertheless, if we suppose that the functions $\phi(y)$ and $\psi(y)$ are continuously differentiable such that

$$\phi'(y) \geq 0 \text{ for all } y \text{ and } \phi'(y) = 0 \text{ for all } y \leq 0, \tag{1.7}$$

then $p(\boldsymbol{x})$ is differentiable, provided that the functions $g_1(\boldsymbol{x}), \ldots, g_m(\boldsymbol{x})$ and $h_1(\boldsymbol{x}), \ldots, h_l(\boldsymbol{x})$ are differentiable. By applying the chain rule, we can write

$$\nabla_{\boldsymbol{x}} p(\boldsymbol{x}) = \sum_{i=1}^{m} \phi'(g_i(\boldsymbol{x}))\, \nabla_{\boldsymbol{x}} g_i(\boldsymbol{x}) + \sum_{i=1}^{l} \psi'(h_i(\boldsymbol{x}))\, \nabla_{\boldsymbol{x}} h_i(\boldsymbol{x}), \tag{1.8}$$

where $\nabla_{\boldsymbol{x}} f(\boldsymbol{x})$ denotes the gradient of a multivariable function $f(\boldsymbol{x}) : \mathbb{R}^n \to \mathbb{R}$ at $\boldsymbol{x} \in \mathbb{R}^n$, i.e.

$$\nabla_{\boldsymbol{x}} f(\boldsymbol{x}) = \left(\frac{\partial f}{\partial x_1}(\boldsymbol{x}), \ldots, \frac{\partial f}{\partial x_n}(\boldsymbol{x}) \right)^T.$$

Let us assume that the assumptions of Theorem 1.1 hold true. Since $\boldsymbol{x}_\mu$ is an optimal solution to the penalty function problem (PP), it is obviously an optimal solution to the following problem

$$\begin{array}{rl} \min & f(\boldsymbol{x}) + \mu p(\boldsymbol{x}) + \mu p_2(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \in \mathbb{R}^n, \end{array} \tag{1.9}$$

where $p_2(\boldsymbol{x}) = \sum_{i=m+1}^{M} \phi(g_i(\boldsymbol{x})) + \sum_{j=l+1}^{L} \psi(h_j(\boldsymbol{x}))$, as $p_2(\boldsymbol{x}) \geq 0$ for all $\boldsymbol{x} \in \mathbb{R}^n$ and the objective function of (1.9) can not be further minimalised.

Since the problem (1.9) is an unconstrained problem and $\boldsymbol{x}_\mu$ is an optimal solution to (1.9), the gradient of the objective function $f(\boldsymbol{x}) + \mu p(\boldsymbol{x}) + \mu p_2(\boldsymbol{x})$ must vanish at $\boldsymbol{x}_\mu$, i.e.

$$\nabla_{\boldsymbol{x}} f(\boldsymbol{x}_\mu) + \mu \nabla_{\boldsymbol{x}} p(\boldsymbol{x}_\mu) + \mu \nabla_{\boldsymbol{x}} p_2(\boldsymbol{x}_\mu) = 0. \tag{1.10}$$

Similarly to (1.8), the gradient of $p_2(\boldsymbol{x}_\mu)$ is as follows

$$\nabla_{\boldsymbol{x}} p_2(\boldsymbol{x}) = \sum_{i=m+1}^{M} \phi'(g_i(\boldsymbol{x})) \nabla_{\boldsymbol{x}} g_i(\boldsymbol{x}) + \sum_{i=l+1}^{L} \psi'(h_i(\boldsymbol{x})) \nabla_{\boldsymbol{x}} h_i(\boldsymbol{x}). \tag{1.11}$$

By substituting the formulas (1.8) and (1.11) into equation (1.10), we obtain

$$\nabla_{\boldsymbol{x}} f(\boldsymbol{x}_\mu) + \sum_{i=1}^{M} \mu\phi'(g_i(\boldsymbol{x}_\mu)) \nabla_{\boldsymbol{x}} g_i(\boldsymbol{x}_\mu) + \sum_{i=1}^{L} \mu\psi'(h_i(\boldsymbol{x}_\mu)) \nabla_{\boldsymbol{x}} h_i(\boldsymbol{x}_\mu) = \mathbf{0}. \tag{1.12}$$

Let $\bar{\boldsymbol{x}}$ be an accumulation point of the sequence $\{\boldsymbol{x}_\mu\}$. Therefore, $\bar{\boldsymbol{x}}$ is an optimal solution to the original problem (OP). Without loss of generality, suppose that the sequence $\{\boldsymbol{x}_\mu\}$ itself converges to $\bar{\boldsymbol{x}}$. Denote $I(\bar{\boldsymbol{x}}) = \{i \in (1,\dots,M) : g_i(\bar{\boldsymbol{x}}) = 0\}$ the set of active inequality constraints at $\bar{\boldsymbol{x}}$, and $N(\bar{\boldsymbol{x}}) = \{i \in (1,\dots,M) : g_i(\bar{\boldsymbol{x}}) < 0\}$ the set of inactive inequality constraints at $\bar{\boldsymbol{x}}$. By the reason of $g_i(\bar{\boldsymbol{x}}) < 0$ for all $i \in N(\bar{\boldsymbol{x}})$, we have $g_i(\boldsymbol{x}_\mu) < 0$ for μ large enough, which gives $\mu\phi'(g_i(\boldsymbol{x}_\mu)) = 0$. Let $\boldsymbol{u}_\mu$ and $\boldsymbol{v}_\mu$ denote vectors having components

$$\begin{array}{ll} (\boldsymbol{u}_\mu)_i = \mu\phi'(g_i(\boldsymbol{x}_\mu)) & \forall\, i \in I(\bar{\boldsymbol{x}}), \\ (\boldsymbol{u}_\mu)_i = 0 & \forall\, i \in N(\bar{\boldsymbol{x}}), \\ (\boldsymbol{v}_\mu)_i = \mu\psi'(h_i(\boldsymbol{x}_\mu)) & \forall\, i = 1,\dots,L. \end{array} \tag{1.13}$$

Since $\mu \geq 0$, and $\psi'(y) \geq 0$, it follows that $(\boldsymbol{u}_\mu)_i \geq 0$ for all $i \in I(\bar{\boldsymbol{x}})$. Having summarised the aforementioned, we can rewrite the foregoing identity (1.12) for all sufficiently large μ as

$$\nabla_{\boldsymbol{x}} f(\boldsymbol{x}_\mu) + \sum_{i=1}^{M} (\boldsymbol{u}_\mu)_i \nabla_{\boldsymbol{x}} g_i(\boldsymbol{x}_\mu) + \sum_{i=1}^{L} (\boldsymbol{v}_\mu)_i \nabla_{\boldsymbol{x}} h_i(\boldsymbol{x}_\mu) = \mathbf{0}. \tag{1.14}$$

Suppose that $\bar{\boldsymbol{x}}$ is a regular solution to the original problem (OP). Then there exist *unique* Lagrangian multipliers $\bar{u}_i \geq 0$, $i \in I(\bar{\boldsymbol{x}})$, $\bar{u}_i = 0$, $i \in N(\bar{\boldsymbol{x}})$, and $\bar{v}_i \in \mathbb{R}$, $i = 1,\dots,L$, such that

$$\nabla_{\boldsymbol{x}} f(\bar{\boldsymbol{x}}) + \sum_{i=1}^{M} \bar{u}_i \nabla_{\boldsymbol{x}} g_i(\bar{\boldsymbol{x}}) + \sum_{i=1}^{L} \bar{v}_i \nabla_{\boldsymbol{x}} h_i(\bar{\boldsymbol{x}}) = \mathbf{0}, \tag{1.15}$$

cf. [3, Theorem 4.3.7]. Denote $\bar{\boldsymbol{u}} = (\bar{u}_1, \ldots, \bar{u}_M)^T$, $\bar{\boldsymbol{v}} = (\bar{v}_1, \ldots, \bar{v}_L)^T$ and suppose that $\left\{\left(\boldsymbol{u}_\mu^T, \boldsymbol{v}_\mu^T\right)^T\right\} \to \left(\bar{\boldsymbol{u}}^T, \bar{\boldsymbol{v}}^T\right)^T$ as $\mu \to \infty$. Since $\boldsymbol{g}(\boldsymbol{x})$, $\boldsymbol{h}(\boldsymbol{x})$, $\phi'(y)$ and $\psi'(y)$ are continuous functions, equations (1.14) converges to (1.15) as $\mu \to \infty$.

It remains to show that $\left\{\left(\boldsymbol{u}_\mu^T, \boldsymbol{v}_\mu^T\right)^T\right\} \to \left(\bar{\boldsymbol{u}}^T, \bar{\boldsymbol{v}}^T\right)^T$ for a unique $\left(\bar{\boldsymbol{u}}^T, \bar{\boldsymbol{v}}^T\right)^T$. Suppose that the sequence $\left\{\left(\boldsymbol{u}_\mu^T, \boldsymbol{v}_\mu^T\right)^T\right\}$ has no accumulation points. Then it holds true that $\left\{\left\|\left(\boldsymbol{u}_\mu^T, \boldsymbol{v}_\mu^T\right)^T\right\|\right\} \to \infty$ as $\mu \to \infty$. Otherwise, the whole sequence would be contained in the bounded set $\{\boldsymbol{z} \in \mathbb{R}^H \colon \|\boldsymbol{z}\| \leq \sup_\mu \left\|\left(\boldsymbol{u}_\mu^T, \boldsymbol{v}_\mu^T\right)^T\right\|\}$, where $H = M + L$. Let us now define a new sequence by

$$\left(\boldsymbol{a}_\mu^T, \boldsymbol{b}_\mu^T\right)^T = \frac{\left(\boldsymbol{u}_\mu^T, \boldsymbol{v}_\mu^T\right)^T}{\left\|\left(\boldsymbol{u}_\mu^T, \boldsymbol{v}_\mu^T\right)^T\right\|}.$$

Since $\left\|\left(\boldsymbol{a}_\mu^T, \boldsymbol{b}_\mu^T\right)^T\right\| = 1$ for all μ, the sequence $\left\{\left(\boldsymbol{a}_\mu^T, \boldsymbol{b}_\mu^T\right)^T\right\}$ is contained in the compact set $\{\boldsymbol{x} \in \mathbb{R}^H \colon \|\boldsymbol{x}\| = 1\}$ and hence has at least one accumulation point. Denote $\left(\bar{\boldsymbol{a}}^T, \bar{\boldsymbol{b}}^T\right)^T$ an accumulation point of the sequence $\left\{\left(\boldsymbol{a}_\mu^T, \boldsymbol{b}_\mu^T\right)^T\right\}$. Without loss of generality, suppose that $\left\{\left(\boldsymbol{a}_\mu^T, \boldsymbol{b}_\mu^T\right)^T\right\}$ itself converges to $\left(\bar{\boldsymbol{a}}^T, \bar{\boldsymbol{b}}^T\right)^T$. Then for sufficiently large μ, and by using equation (1.14), the following identities hold true

$$\begin{aligned}
&\sum_{i \in I(\bar{\boldsymbol{x}})} (\boldsymbol{a}_\mu)_i \nabla_{\boldsymbol{x}} g_i(\boldsymbol{x}_\mu) + \sum_{i=1}^{L} (\boldsymbol{b}_\mu)_i \nabla_{\boldsymbol{x}} h_i(\boldsymbol{x}_\mu) \\
&= \frac{1}{\left\|\left(\boldsymbol{u}_\mu^T, \boldsymbol{v}_\mu^T\right)^T\right\|} \left(\sum_{i \in I(\bar{\boldsymbol{x}})} (\boldsymbol{u}_\mu)_i \nabla_{\boldsymbol{x}} g_i(\boldsymbol{x}_\mu) + \sum_{i=1}^{L} (\boldsymbol{v}_\mu)_i \nabla_{\boldsymbol{x}} h_i(\boldsymbol{x}_\mu)\right) \\
&= \frac{1}{\left\|\left(\boldsymbol{u}_\mu^T, \boldsymbol{v}_\mu^T\right)^T\right\|} \left(\sum_{i=1}^{M} (\boldsymbol{u}_\mu)_i \nabla_{\boldsymbol{x}} g_i(\boldsymbol{x}_\mu) + \sum_{i=1}^{L} (\boldsymbol{v}_\mu)_i \nabla_{\boldsymbol{x}} h_i(\boldsymbol{x}_\mu)\right) \\
&= -\frac{\nabla_{\boldsymbol{x}} f(\boldsymbol{x}_\mu)}{\left\|\left(\boldsymbol{u}_\mu^T, \boldsymbol{v}_\mu^T\right)^T\right\|}.
\end{aligned} \tag{1.16}$$

As $\mu \to \infty$, we have $\{\boldsymbol{x}\}_\mu \to \bar{\boldsymbol{x}}$, $\left\{\left(\boldsymbol{a}_\mu^T, \boldsymbol{b}_\mu^T\right)^T\right\} \to \left(\bar{\boldsymbol{a}}^T, \bar{\boldsymbol{b}}^T\right)^T$, and $\left\{\left\|\left(\boldsymbol{u}_\mu^T, \boldsymbol{v}_\mu^T\right)^T\right\|\right\} \to \infty$, and therefore equation (1.16) becomes

$$\sum_{i \in I(\bar{\boldsymbol{x}})} (\bar{\boldsymbol{a}})_i \nabla_{\boldsymbol{x}} g_i(\bar{\boldsymbol{x}}) + \sum_{i=1}^{L} (\bar{\boldsymbol{b}})_i \nabla_{\boldsymbol{x}} h_i(\bar{\boldsymbol{x}}) = \boldsymbol{0}. \tag{1.17}$$

Since $\left\|\left(\bar{\boldsymbol{a}}^T, \bar{\boldsymbol{b}}^T\right)^T\right\| = 1$, equation (1.17) contradicts the linear independence of vectors $\nabla_{\boldsymbol{x}} g_i(\bar{\boldsymbol{x}})$, $i \in I(\bar{\boldsymbol{x}})$, and $\nabla_{\boldsymbol{x}} h_1(\bar{\boldsymbol{x}}), \ldots, \nabla_{\boldsymbol{x}} h_L(\bar{\boldsymbol{x}})$. Hence, $\left\{\left(\boldsymbol{u}_\mu^T, \boldsymbol{v}_\mu^T\right)^T\right\}$ has at least one accumulation point.

To show its uniqueness, suppose that $\left\{\left(\boldsymbol{u}_{\mu}^{T},\boldsymbol{v}_{\mu}^{T}\right)^{T}\right\}$ has two accumulation points, $\left(\bar{\boldsymbol{u}}^{T},\bar{\boldsymbol{v}}^{T}\right)^{T}$ and $\left(\tilde{\boldsymbol{u}}^{T},\tilde{\boldsymbol{v}}^{T}\right)^{T}$. Having noted that $\bar{u}_i = 0$ and $\tilde{u}_i = 0$ for $i \in N(\bar{\boldsymbol{x}})$ necessarily, we can write

$$\sum_{i\in I(\bar{\boldsymbol{x}})} \bar{u}_i \nabla_{\boldsymbol{x}} g_i(\bar{\boldsymbol{x}}) + \sum_{i=1}^{L} \bar{v}_i \nabla_{\boldsymbol{x}} h_i(\bar{\boldsymbol{x}}) = -\nabla_{\boldsymbol{x}} f(\bar{\boldsymbol{x}}) = \sum_{i\in I(\bar{\boldsymbol{x}})} \tilde{u}_i \nabla_{\boldsymbol{x}} g_i(\bar{\boldsymbol{x}}) + \sum_{i=1}^{L} \tilde{v}_i \nabla_{\boldsymbol{x}} h_i(\bar{\boldsymbol{x}}),$$

which yields

$$\sum_{i\in I(\bar{\boldsymbol{x}})} (\bar{u}_i - \tilde{u}_i) \nabla_{\boldsymbol{x}} g_i(\bar{\boldsymbol{x}}) + \sum_{i=1}^{L} (\bar{v}_i - \tilde{v}_i) \nabla_{\boldsymbol{x}} h_i(\bar{\boldsymbol{x}}) = \boldsymbol{0}.$$

By considering the assumption of linear independence of the gradients, we obtain $\bar{u}_i - \tilde{u}_i = 0$ for all $i \in I(\bar{\boldsymbol{x}})$, and $\bar{v}_i - \tilde{v}_i = 0$ for all $i = 1, \ldots, L$. Thus, we showed that $\left(\bar{\boldsymbol{u}}^{T},\bar{\boldsymbol{v}}^{T}\right)^{T} = \left(\tilde{\boldsymbol{u}}^{T},\tilde{\boldsymbol{v}}^{T}\right)^{T}$.

For the sake of perspicuity, let us formulate the above stated facts in a proposition.

Propositon 1.1. *Suppose that $\phi(y)$ and $\psi(y)$ are continuously differentiable functions, and $\phi(y)$ satisfies*

$$\phi'(y) \geq 0 \text{ for all } y \text{ and } \phi'(y) = 0 \text{ for all } y \leq 0.$$

Futhermore, assume that

$$\mathcal{X} = \{\boldsymbol{x} \in \mathbb{R}^n \colon g_i(\boldsymbol{x}) \leq 0,\ i = m+1, \ldots, M,\ h_i(\boldsymbol{x}) = 0,\ i = l+1, \ldots, L\}, \tag{1.18}$$

$f(\boldsymbol{x})$, $g_1(\boldsymbol{x}), \ldots, g_M(\boldsymbol{x})$ and $h_1(\boldsymbol{x}), \ldots, h_L(\boldsymbol{x})$ are differentiable. Moreover, suppose that the assumptions of Theorem 1.1 hold true, and $\bar{\boldsymbol{x}}$ is a regular solution to the original problem (OP)*. Denote $I(\bar{\boldsymbol{x}}) = \{i \in (1, \ldots, M) : g_i(\bar{\boldsymbol{x}}) = 0\}$ the set of active inequality constraints at $\bar{\boldsymbol{x}}$, and $N(\bar{\boldsymbol{x}}) = \{i \in (1, \ldots, M) : g_i(\bar{\boldsymbol{x}}) < 0\}$ the set of inactive inequality constraints at $\bar{\boldsymbol{x}}$. Define $\left\{\left(\boldsymbol{u}_{\mu}^{T},\boldsymbol{v}_{\mu}^{T}\right)^{T}\right\}$ by*

$$\begin{aligned}
(\boldsymbol{u}_{\mu})_i &= \mu\phi'(g_i(\boldsymbol{x}_{\mu})) \qquad &&\forall\ i \in I(\bar{\boldsymbol{x}}),\\
(\boldsymbol{u}_{\mu})_i &= 0 &&\forall\ i \in N(\bar{\boldsymbol{x}}),\\
(\boldsymbol{v}_{\mu})_i &= \mu\psi'(h_i(\boldsymbol{x}_{\mu})) &&\forall\ i = 1, \ldots, L.
\end{aligned}$$

Then $\left\{\left(\boldsymbol{u}_{\mu}^{T},\boldsymbol{v}_{\mu}^{T}\right)^{T}\right\} \to \left(\bar{\boldsymbol{u}}^{T},\bar{\boldsymbol{v}}^{T}\right)^{T}$ as $\mu \to \infty$, where $\left(\bar{\boldsymbol{u}}^{T},\bar{\boldsymbol{v}}^{T}\right)^{T}$ is the vector of KKT Lagrange multipliers associated with the inequality and equality constraints at optimality.

That is to say, Proposition 1.1 states that $\boldsymbol{u}_{\mu}$ and $\boldsymbol{v}_{\mu}$ can be used as estimates of the KKT Lagrange multipliers at optimality for sufficiently large μ.

Remark that if $\bar{u}_i > 0$ for some $i \in I(\bar{\boldsymbol{x}})$, then $(\boldsymbol{u}_{\mu})_i > 0$ for sufficiently large μ. This means that $g_i(\boldsymbol{x}) \leq 0$ is violated along the trajectory leading to $\bar{\boldsymbol{x}}$. On the other hand, observe that $g_i(\boldsymbol{x}_{\mu}) \leq 0$ for all $i = m+1, \ldots, M$, therefore $\bar{u}_i = 0$. In other words, if a solution to the original problem (OP) for sets $\mathcal{X}$ of type (1.18) satisfies the *strict complementary slackness condition* ([3, p. 499]) then $I(\bar{\boldsymbol{x}}) \cap \{m+1, \ldots, M\} = \emptyset$.

1.2.2 An example of exterior penalty function

The quadratic penalty function

The quadratic penalty function for (OP) is a special case of L_q penalty function (1.5) for $q = 2$. Therefore, it is given by

$$p_2(\boldsymbol{x}) = \| \max\{\mathbf{0}, \boldsymbol{g}(\boldsymbol{x})\}\|_2^2 + \|\boldsymbol{h}(\boldsymbol{x})\|_2^2 = \sum_{i=1}^{m}[\max\{0, g_i(\boldsymbol{x})\}]^2 + \sum_{j=1}^{l} |h_j(\boldsymbol{x})|^2. \tag{1.19}$$

Let us demonstrate, that the quadratic penalty function satisfies (1.13), and thus it can recover the KKT Lagrange multipliers. From the definition of the quadratic penalty function it follows that

$$\phi_2(y) = \max\{0, y\}^2 \qquad \text{and} \qquad \psi_2(y) = y^2.$$

Having differentiated them, we can write

$$\phi_2'(y) = 2\max\{0, y\} \qquad \text{and} \qquad \psi_2'(y) = 2y.$$

Evidently, $\phi_2'(y) \geq 0$ for all y, and $\phi_2'(y) = 0$ for $y \leq 0$. Therefore, we easily obtain from Proposition 1.1 that

$$\begin{aligned} (\boldsymbol{u}_\mu)_i &= 2\mu \max\{0, g_i(\boldsymbol{x}_\mu)\} \quad \forall\, i \in I(\bar{\boldsymbol{x}}), \\ (\boldsymbol{v}_\mu)_i &= 2\mu h_i(\boldsymbol{x}_\mu) \quad \forall\, i = 1, \ldots, L. \end{aligned}$$

1.3 Exact penalty function methods

In this section, we shall associate with the original problem (OP) the following class of penalty functions

$$p(\boldsymbol{x}) = \|\max\{\mathbf{0}, \boldsymbol{g}(\boldsymbol{x})\}\| + \|\boldsymbol{h}(\boldsymbol{x})\|, \tag{1.20}$$

where $\|\cdot\|$ is any fixed vector norm in $\mathbb{R}^m$ and in $\mathbb{R}^l$, respectively. A similar class of penalty function was put forward in [17]. In this section we denote the objective function of the penalty problem (PP) as $F_E(\boldsymbol{x}; \mu)$. Thus, we have

$$F_E(\boldsymbol{x}; \mu) = f(\boldsymbol{x}) + \mu p(\boldsymbol{x}) = f(\boldsymbol{x}) + \mu\|\max\{\mathbf{0}, \boldsymbol{g}(\boldsymbol{x})\}\| + \mu\|\boldsymbol{h}(\boldsymbol{x})\|.$$

On the account of the presence of the norm and the function $\max\{\mathbf{0}, \cdot\}$, $p(\boldsymbol{x})$ is non-differentiable at some $\boldsymbol{x}$.

The exact penalty method form of the original problem (OP) is formulated as follows

$$\begin{aligned} \min \quad & F_E(\boldsymbol{x}; \mu) \\ \text{subject to} \quad & \boldsymbol{x} \in \mathcal{X}. \end{aligned} \tag{EP}$$

The following theorem establishes the exactness of the (1.20) penalty function. Its proof is partially based on the steps of the proofs of Theorems 3 and 5 in [20].

Theorem 1.2. *Suppose that $\bar{\boldsymbol{x}} \in S$ is a KKT point for the original problem* (OP) *with Lagrangian multipliers $\bar{u}_i$, $i = 1, \ldots, m$, and $\bar{v}_i$, $i = 1, \ldots, l$ associated with the inequality $g_i(\boldsymbol{x})$, $i = 1, \ldots, m$, and the equality constraints $h_i(\boldsymbol{x})$, $i = 1, \ldots, l$,* *respectively. Moreover, suppose that the second-order sufficient conditions hold true at* $\bar{\boldsymbol{x}}$ *and*

$$\mu > \bar{\mu} = \max\left\{\|\bar{\boldsymbol{u}}\|', \|\bar{\boldsymbol{v}}\|'\right\},$$

where $\|\cdot\|'$ *is the dual norm to* $\|\cdot\|$, $\bar{\boldsymbol{u}} = (\bar{u}_1, \ldots, \bar{u}_m)^T$, *and* $\bar{\boldsymbol{v}} = (\bar{v}_1, \ldots, \bar{v}_l)^T$.

Then $\bar{\boldsymbol{x}}$ *is a strict local solution to the exact penalty function problem* (EP) *for all* $\mu > \bar{\mu}$.

Proof. To show that $\bar{\boldsymbol{x}}$ is a strict local solution to (EP), we have to solve the following auxiliary problem.

Without loss of generality, suppose that

$$\mathcal{X} = \{\boldsymbol{x} \in \mathbb{R}^n \colon g_i(\boldsymbol{x}) \leq 0,\ i = m+1, \ldots, M,\ h_i(\boldsymbol{x}) = 0,\ i = l+1, \ldots, L\}.$$

Furthermore, let $\bar{u}_i^2$, $i = m+1, \ldots, M$, and $\bar{v}_i^2$, $i = l+1, \ldots, L$ be the Lagrangian multipliers associated with $g_i(\bar{\boldsymbol{x}})$, $i = m+1, \ldots, M$, and $h_i(\bar{\boldsymbol{x}})$, $i = l+1, \ldots, L$ and denote $\boldsymbol{g_2}(\boldsymbol{x}) = (g_{m+1}(\boldsymbol{x}), \ldots, g_M(\boldsymbol{x}))^T$, $\boldsymbol{h_2}(\boldsymbol{x}) = (h_{l+1}(\boldsymbol{x}), \ldots, h_L(\boldsymbol{x}))^T$, $\bar{\boldsymbol{u}}_{\boldsymbol{2}} = \left(\bar{u}_{m+1}^2, \ldots, \bar{u}_M^2\right)^T$, and $\bar{\boldsymbol{v}}_{\boldsymbol{2}} = \left(\bar{v}_{l+1}^2, \ldots, \bar{v}_L^2\right)^T$. Let $\|\cdot\|$ be any fixed vector norm in $\mathbb{R}^{(M-m+L-l)}$. Then the auxiliary problem is formulated as follows

$$\begin{array}{rl} \min & F_{AE}(\boldsymbol{x}; \mu, \mu_2) \\ \text{subject to} & \boldsymbol{x} \in \mathbb{R}^n, \end{array} \tag{AEP}$$

where $F_{AE}(\boldsymbol{x}; \mu, \mu_2) = F_E(\boldsymbol{x}; \mu) + \mu_2 \left\|\left(\max\left\{\boldsymbol{0}, \boldsymbol{g_2}(\boldsymbol{x})\right\}^T, \boldsymbol{h_2}(\boldsymbol{x})^T\right)^T\right\|$. Suppose that $\mu_2 > \bar{\mu}_2 = \left\|\left(\bar{\boldsymbol{u}}_{\boldsymbol{2}}^T, \bar{\boldsymbol{v}}_{\boldsymbol{2}}^T\right)^T\right\|'$. We will show that $\bar{\boldsymbol{x}}$ is a strict local solution to (AEP).

Assume that $\bar{\boldsymbol{x}}$ is not a strict local solution to (AEP). Thus, there exists a sequence $\left\{\boldsymbol{x}^k\right\}$, $\boldsymbol{x}^k \neq \bar{\boldsymbol{x}}$, converging to $\bar{\boldsymbol{x}}$ such that $F_{AE}(\boldsymbol{x}^k; \mu, \mu_2) \leq F_{AE}(\bar{\boldsymbol{x}}; \mu, \mu_2)$ for all $k \geq 0$. Thereby,

$$\begin{aligned} F_{AE}\left(x^k; \mu, \mu_2\right) - F_{AE}(\bar{\boldsymbol{x}}; \mu, \mu_2) &= f(\boldsymbol{x}^k) - f(\bar{\boldsymbol{x}}) + \mu \left\|\max\left\{\boldsymbol{0}, \boldsymbol{g}\left(\boldsymbol{x}^k\right)\right\}\right\| \\ &+ \mu \left\|\boldsymbol{h}\left(\boldsymbol{x}^k\right)\right\| + \mu_2 \left\|\left(\max\left\{\boldsymbol{0}, \boldsymbol{g_2}(\boldsymbol{x})\right\}^T, \boldsymbol{h_2}(\boldsymbol{x})^T\right)^T\right\| \leq 0 \end{aligned} \tag{1.21}$$

Having noted that the set $\{\boldsymbol{x} \in \mathbb{R}^n \colon \|x\| = 1\}$ is compact, we define the sequence $\{\boldsymbol{d}^k\}$ as

$$\boldsymbol{d}^k = \frac{\boldsymbol{x}^k - \bar{\boldsymbol{x}}}{\|\boldsymbol{x}^k - \bar{\boldsymbol{x}}\|}$$

and take a convergent subsequence $\{\boldsymbol{d}^{k_l}\}$ that converges to $\boldsymbol{d}$. After dividing (1.21) by $\|\boldsymbol{x}^{k_l} - \bar{\boldsymbol{x}}\|$ and subtracting $\boldsymbol{g}(\bar{\boldsymbol{x}}) = \boldsymbol{0}$, $\boldsymbol{h}(\bar{\boldsymbol{x}}) = \boldsymbol{0}$, $\boldsymbol{g_2}(\bar{\boldsymbol{x}}) = \boldsymbol{0}$,$\boldsymbol{h_2}(\bar{\boldsymbol{x}}) = \boldsymbol{0}$, we get

$$\begin{aligned} &\frac{f\left(\boldsymbol{x}^{k_l}\right) - f(\bar{\boldsymbol{x}})}{\|\boldsymbol{x}^{k_l} - \bar{\boldsymbol{x}}\|} + \mu \left\|\max\left\{\boldsymbol{0}, \frac{\boldsymbol{g}\left(\boldsymbol{x}^{k_l}\right) - \boldsymbol{g}(\bar{\boldsymbol{x}})}{\|\boldsymbol{x}^{k_l} - \bar{\boldsymbol{x}}\|}\right\}\right\| + \mu \left\|\frac{\boldsymbol{h}\left(\boldsymbol{x}^{k_l}\right) - \boldsymbol{h}(\bar{\boldsymbol{x}})}{\|\boldsymbol{x}^{k_l} - \bar{\boldsymbol{x}}\|}\right\| \\ &+ \mu_2 \left\|\left(\max\left\{\boldsymbol{0}, \frac{\boldsymbol{g_2}\left(\boldsymbol{x}^{k_l}\right) - \boldsymbol{g_2}(\bar{\boldsymbol{x}})}{\|\boldsymbol{x}^{k_l} - \bar{\boldsymbol{x}}\|}\right\}^T, \left(\frac{\boldsymbol{h_2}\left(\boldsymbol{x}^{k_l}\right) - \boldsymbol{h_2}(\bar{\boldsymbol{x}})}{\|\boldsymbol{x}^{k_l} - \bar{\boldsymbol{x}}\|}\right)^T\right)^T\right\| \leq 0. \end{aligned} \tag{1.22}$$

Note that f, g_i, $i = 1, \ldots, M$, h_i, $i = 1, \ldots, L$ are differentiable at $\bar{\boldsymbol{x}}$, $\|\cdot\|$ and $\max\{\mathbf{0}, \cdot\}$ are continuous. Then by taking the limit of (1.22) as k_l approaches infinity, we obtain

$$\begin{aligned} &\nabla_{\boldsymbol{x}} f(\bar{\boldsymbol{x}})^T \boldsymbol{d} + \mu \left\| J_{g^+}(\bar{\boldsymbol{x}})\, \boldsymbol{d} \right\| + \mu \left\| J_h(\bar{\boldsymbol{x}})\, \boldsymbol{d} \right\| \\ &+\mu_2 \left\| \left(\left(J_{g_2^+}(\bar{\boldsymbol{x}})\, \boldsymbol{d} \right)^T, \left(J_{h_2}(\bar{\boldsymbol{x}})\, \boldsymbol{d} \right)^T \right)^T \right\| \leq 0, \end{aligned} \tag{1.23}$$

where

$$\begin{aligned} J_{g^+}(\bar{\boldsymbol{x}})\boldsymbol{d} &= \left(\max\left\{ \nabla_{\boldsymbol{x}} g_1(\bar{\boldsymbol{x}})^T \boldsymbol{d}, 0 \right\}, \ldots, \max\left\{ \nabla_{\boldsymbol{x}} g_m(\bar{\boldsymbol{x}})^T \boldsymbol{d}, 0 \right\} \right)^T, \\ J_h(\bar{\boldsymbol{x}})\boldsymbol{d} &= \left(\nabla_{\boldsymbol{x}} h_1(\bar{\boldsymbol{x}})^T \boldsymbol{d}, \ldots, \nabla_{\boldsymbol{x}} h_l(\bar{\boldsymbol{x}})^T \boldsymbol{d} \right)^T \end{aligned}$$

and

$$\begin{aligned} J_{g_2^+}(\bar{\boldsymbol{x}})\boldsymbol{d} &= \left(\max\left\{ \nabla_{\boldsymbol{x}} g_{m+1}(\bar{\boldsymbol{x}})^T \boldsymbol{d}, 0 \right\}, \ldots, \max\left\{ \nabla_{\boldsymbol{x}} g_M(\bar{\boldsymbol{x}})^T \boldsymbol{d}, 0 \right\} \right)^T, \\ J_{h_2}(\bar{\boldsymbol{x}})\boldsymbol{d} &= \left(\nabla_{\boldsymbol{x}} h_{l+1}(\bar{\boldsymbol{x}})^T \boldsymbol{d}, \ldots, \nabla_{\boldsymbol{x}} h_L(\bar{\boldsymbol{x}})^T \boldsymbol{d} \right)^T. \end{aligned}$$

Since $\bar{\boldsymbol{x}}$ is a KKT point for (OP), therefore

$$\nabla_{\boldsymbol{x}} f(\bar{\boldsymbol{x}})^T \boldsymbol{d} = -\bar{\boldsymbol{u}}^T J_g(\bar{\boldsymbol{x}}) \boldsymbol{d} - \bar{\boldsymbol{v}}^T J_h(\bar{\boldsymbol{x}}) \boldsymbol{d} - \bar{\boldsymbol{u}}_{\mathbf{2}}^T J_{g_2}(\bar{\boldsymbol{x}}) - \bar{\boldsymbol{v}}_{\mathbf{2}}^T J_{h_2}(\bar{\boldsymbol{x}}) \boldsymbol{d}. \tag{1.24}$$

By using (1.24) in (1.23), we find that

$$\begin{aligned} &\mu \left\| J_{g^+}(\bar{\boldsymbol{x}})\, \boldsymbol{d} \right\| + \mu \left\| J_h(\bar{\boldsymbol{x}})\, \boldsymbol{d} \right\| - \bar{\boldsymbol{u}}^T J_g(\bar{\boldsymbol{x}}) \boldsymbol{d} - \bar{\boldsymbol{v}}^T J_h(\bar{\boldsymbol{x}}) \boldsymbol{d} \\ &+\mu_2 \left\| \left(\left(J_{g_2^+}(\bar{\boldsymbol{x}})\, \boldsymbol{d} \right)^T, \left(J_{h_2}(\bar{\boldsymbol{x}})\, \boldsymbol{d} \right)^T \right)^T \right\| - \bar{\boldsymbol{u}}_{\mathbf{2}}^T J_{g_2}(\bar{\boldsymbol{x}}) \boldsymbol{d} - \bar{\boldsymbol{v}}_{\mathbf{2}}^T J_{h_2}(\bar{\boldsymbol{x}}) \boldsymbol{d} \leq 0. \end{aligned} \tag{1.25}$$

By noting that $J_g(\bar{\boldsymbol{x}})\boldsymbol{d} \leq J_{g^+}(\bar{\boldsymbol{x}})\boldsymbol{d}$, and $J_{g_2}(\bar{\boldsymbol{x}})\boldsymbol{d} \leq J_{g_2^+}(\bar{\boldsymbol{x}})\boldsymbol{d}$, we get

$$\begin{aligned} &\mu \left\| J_{g^+}(\bar{\boldsymbol{x}})\, \boldsymbol{d} \right\| + \mu \left\| J_h(\bar{\boldsymbol{x}})\, \boldsymbol{d} \right\| - \bar{\boldsymbol{u}}^T J_{g^+}(\bar{\boldsymbol{x}}) \boldsymbol{d} - \bar{\boldsymbol{v}}^T J_h(\bar{\boldsymbol{x}}) \boldsymbol{d} \\ &+\mu_2 \left\| \left(\left(J_{g_2^+}(\bar{\boldsymbol{x}})\, \boldsymbol{d} \right)^T, \left(J_{h_2}(\bar{\boldsymbol{x}})\, \boldsymbol{d} \right)^T \right)^T \right\| - \bar{\boldsymbol{u}}_{\mathbf{2}}^T J_{g_2^+}(\bar{\boldsymbol{x}}) \boldsymbol{d} - \bar{\boldsymbol{v}}_{\mathbf{2}}^T J_{h_2}(\bar{\boldsymbol{x}}) \boldsymbol{d} \leq 0. \end{aligned}$$

By using the generalised Cauchy–Schwarz inequality (A.1), we conclude that

$$\begin{aligned} &\left(\mu - \|\bar{\boldsymbol{u}}\|' \right) \left\| J_{g^+}(\bar{\boldsymbol{x}}) \boldsymbol{d} \right\| + \left(\mu - \|\bar{\boldsymbol{v}}\|' \right) \left\| J_h(\bar{\boldsymbol{x}})\, \boldsymbol{d} \right\| \\ &+ \left(\mu_2 - \left\| \left(\bar{\boldsymbol{u}}_{\mathbf{2}}^T, \bar{\boldsymbol{v}}_{\mathbf{2}}^T \right)^T \right\|' \right) \left\| \left(\left(J_{g_2^+}(\bar{\boldsymbol{x}})\, \boldsymbol{d} \right)^T, \left(J_{h_2}(\bar{\boldsymbol{x}})\, \boldsymbol{d} \right)^T \right)^T \right\| \leq 0 \end{aligned} \tag{1.26}$$

Since $\mu > \max\left\{ \|\bar{\boldsymbol{u}}\|', \|\bar{\boldsymbol{v}}\|' \right\}$, and $\mu_2 > \left\| \left(\bar{\boldsymbol{u}}_{\mathbf{2}}^T, \bar{\boldsymbol{v}}_{\mathbf{2}}^T \right)^T \right\|'$, it follows from (1.26) that

$$J_{g^+}(\bar{\boldsymbol{x}})\boldsymbol{d} = \mathbf{0}, \;\; J_h(\bar{\boldsymbol{x}})\boldsymbol{d} = \mathbf{0}, \;\; J_{g_2^+}(\bar{\boldsymbol{x}})\boldsymbol{d} = \mathbf{0}, \;\text{ and } J_{h_2}(\bar{\boldsymbol{x}})\boldsymbol{d} = \mathbf{0}. \tag{1.27}$$

Denote

$$I(\bar{\boldsymbol{x}}) = \left\{ i \in (1, \ldots, m) : g_i(\bar{\boldsymbol{x}}) = 0 \right\},$$

and

$$I_2(\bar{\boldsymbol{x}}) = \left\{ i \in (m+1, \ldots, M) : g_i(\bar{\boldsymbol{x}}) = 0 \right\}.$$

From [3, Theorem 4.4.2], we know that the following must hold true

$$\begin{aligned}
\nabla_{\boldsymbol{x}} g_i(\bar{\boldsymbol{x}})^T \boldsymbol{d} \leq 0 \quad & \text{for} \quad i \in I\left(\bar{\boldsymbol{x}}\right), \\
\nabla_{\boldsymbol{x}} g_i(\bar{\boldsymbol{x}})^T \boldsymbol{d} \leq 0 \quad & \text{for} \quad i \in I_2\left(\bar{\boldsymbol{x}}\right), \\
\nabla_{\boldsymbol{x}} h_i(\bar{\boldsymbol{x}})^T \boldsymbol{d} = 0 \quad & \text{for} \quad i = 1, \ldots, L.
\end{aligned} \tag{1.28}$$

By using identities (1.27) in (1.25), we obtain

$$\begin{aligned}
\nabla_{\boldsymbol{x}} g_i(\bar{\boldsymbol{x}})^T \boldsymbol{d} \geq 0 \quad & \text{for} \quad i \in I\left(\bar{\boldsymbol{x}}\right), \bar{u}_i > 0, \\
\nabla_{\boldsymbol{x}} g_i(\bar{\boldsymbol{x}})^T \boldsymbol{d} \geq 0 \quad & \text{for} \quad i \in I_2\left(\bar{\boldsymbol{x}}\right), \bar{u}_i^2 > 0, .
\end{aligned} \tag{1.29}$$

Thus, the following identities hold true

$$\begin{aligned}
\nabla_{\boldsymbol{x}} g_i(\bar{\boldsymbol{x}})^T \boldsymbol{d} = 0 \quad & \text{for} \quad i \in I\left(\bar{\boldsymbol{x}}\right), \bar{u}_i > 0, \\
\nabla_{\boldsymbol{x}} g_i(\bar{\boldsymbol{x}})^T \boldsymbol{d} \leq 0 \quad & \text{for} \quad i \in I\left(\bar{\boldsymbol{x}}\right), \bar{u}_i = 0, \\
\nabla_{\boldsymbol{x}} g_i(\bar{\boldsymbol{x}})^T \boldsymbol{d} = 0 \quad & \text{for} \quad i \in I_2\left(\bar{\boldsymbol{x}}\right), \bar{u}_i^2 > 0, \\
\nabla_{\boldsymbol{x}} g_i(\bar{\boldsymbol{x}})^T \boldsymbol{d} \leq 0 \quad & \text{for} \quad i \in I_2\left(\bar{\boldsymbol{x}}\right), \bar{u}_i^2 = 0, \\
\nabla_{\boldsymbol{x}} h_i(\bar{\boldsymbol{x}})^T \boldsymbol{d} = 0 \quad & \text{for} \quad i = 1, \ldots, L.
\end{aligned} \tag{1.30}$$

Identities (1.30) imply that $\boldsymbol{d}$ is in the critical cone, viz. (A.2). After expanding $\mathcal{L}\left(\cdot; \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}, \bar{\boldsymbol{u}}_{\boldsymbol{2}}, \bar{\boldsymbol{v}}_{\boldsymbol{2}}\right)$ in $\bar{\boldsymbol{x}}$, we can write

$$\begin{aligned}
\mathcal{L}\left(\boldsymbol{x}^{k_l}; \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}, \bar{\boldsymbol{u}}_{\boldsymbol{2}}, \bar{\boldsymbol{v}}_{\boldsymbol{2}}\right) - \mathcal{L}\left(\bar{\boldsymbol{x}}; \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}, \bar{\boldsymbol{u}}_{\boldsymbol{2}}, \bar{\boldsymbol{v}}_{\boldsymbol{2}}\right) = \nabla_{\boldsymbol{x}} \mathcal{L}\left(\bar{\boldsymbol{x}}; \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}, \bar{\boldsymbol{u}}_{\boldsymbol{2}}, \bar{\boldsymbol{v}}_{\boldsymbol{2}}\right)^T \left(\boldsymbol{x}^{k_l} - \bar{\boldsymbol{x}}\right) \\
+ \frac{1}{2}\left(\boldsymbol{x}^{k_l} - \bar{\boldsymbol{x}}\right)^T \nabla^2_{\boldsymbol{xx}} \mathcal{L}\left(\bar{\boldsymbol{x}}; \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}, \bar{\boldsymbol{u}}_{\boldsymbol{2}}, \bar{\boldsymbol{v}}_{\boldsymbol{2}}\right)\left(\boldsymbol{x}^{k_l} - \bar{\boldsymbol{x}}\right) + \left\|\boldsymbol{x}^{k_l} - \bar{\boldsymbol{x}}\right\|^2 \alpha_{\mathcal{L}}\left(\bar{\boldsymbol{x}}; \boldsymbol{x}^{k_l} - \bar{\boldsymbol{x}}\right),
\end{aligned} \tag{1.31}$$

where $\lim_{k_l \to \infty} \alpha_{\mathcal{L}}\left(\bar{\boldsymbol{x}}; \boldsymbol{x}^{k_l} - \bar{\boldsymbol{x}}\right) = 0$, vide [3, definition 3.3.5]. As $\bar{\boldsymbol{x}}$ is a KKT point, $\nabla_{\boldsymbol{x}} \mathcal{L}\left(\bar{\boldsymbol{x}}, \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}, \bar{\boldsymbol{u}}_{\boldsymbol{2}}, \bar{\boldsymbol{v}}_{\boldsymbol{2}}\right) = \boldsymbol{0}$. Since the second-order sufficient conditions hold true, $\boldsymbol{d}$ is in the critical cone and

$$\begin{aligned}
\lim_{k_l \to \infty} \frac{1}{2} \frac{\left(\boldsymbol{x}^{k_l} - \bar{\boldsymbol{x}}\right)^T}{\left\|\boldsymbol{x}^{k_l} - \bar{\boldsymbol{x}}\right\|} \nabla^2_{\boldsymbol{xx}} \mathcal{L}\left(\bar{\boldsymbol{x}}; \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}, \bar{\boldsymbol{u}}_{\boldsymbol{2}}, \bar{\boldsymbol{v}}_{\boldsymbol{2}}\right) \frac{\left(\boldsymbol{x}^{k_l} - \bar{\boldsymbol{x}}\right)}{\left\|\boldsymbol{x}^{k_l} - \bar{\boldsymbol{x}}\right\|}. \\
+ \alpha_{\mathcal{L}}\left(\bar{\boldsymbol{x}}; \boldsymbol{x}^{k_l} - \bar{\boldsymbol{x}}\right) = \frac{1}{2} \boldsymbol{d}^T \nabla^2_{\boldsymbol{xx}} \mathcal{L}\left(\bar{\boldsymbol{x}}; \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}, \bar{\boldsymbol{u}}_{\boldsymbol{2}}, \bar{\boldsymbol{v}}_{\boldsymbol{2}}\right) \boldsymbol{d} > 0,
\end{aligned}$$

and hence for sufficiently large k_l

$$\frac{1}{2}\left(\boldsymbol{x}^{k_l} - \bar{\boldsymbol{x}}\right)^T \nabla^2_{\boldsymbol{xx}} \mathcal{L}\left(\bar{\boldsymbol{x}}; \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}, \bar{\boldsymbol{u}}_{\boldsymbol{2}}, \bar{\boldsymbol{v}}_{\boldsymbol{2}}\right)\left(\boldsymbol{x}^{k_l} - \bar{\boldsymbol{x}}\right) + \left\|\boldsymbol{x}^{k_l} - \bar{\boldsymbol{x}}\right\|^2 \alpha_{\mathcal{L}}\left(\bar{\boldsymbol{x}}; \boldsymbol{x}^{k_l} - \bar{\boldsymbol{x}}\right) > 0. \tag{1.32}$$

By using (1.32) in (1.31), for sufficiently large k_l, we get

$$\mathcal{L}\left(\boldsymbol{x}^{k_l}; \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}, \bar{\boldsymbol{u}}_{\boldsymbol{2}}, \bar{\boldsymbol{v}}_{\boldsymbol{2}}\right) > \mathcal{L}\left(\bar{\boldsymbol{x}}; \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}, \bar{\boldsymbol{u}}_{\boldsymbol{2}}, \bar{\boldsymbol{v}}_{\boldsymbol{2}}\right). \tag{1.33}$$

By using the generalised Cauchy–Schwarz inequality (A.1), we obtain

$$\begin{aligned}
F_{AE}\left(\boldsymbol{x}^{k_l}; \mu, \mu_2\right) = {} & f\left(\boldsymbol{x}^{k_l}\right) + \mu \left\|\max\left\{\boldsymbol{0}, \boldsymbol{g}\left(\boldsymbol{x}^{k_l}\right)\right\}\right\| + \mu \left\|\boldsymbol{h}\left(\boldsymbol{x}^{k_l}\right)\right\| \\
& + \mu_2 \left\|\left(\max\left\{\boldsymbol{0}, \boldsymbol{g_2}\left(\boldsymbol{x}^{k_l}\right)\right\}^T, \boldsymbol{h_2}\left(\boldsymbol{x}^{k_l}\right)^T\right)^T\right\| \\
& > f\left(\boldsymbol{x}^{k_l}\right) + \left\|\bar{\boldsymbol{u}}\right\|' \left\|\max\left\{\boldsymbol{0}, \boldsymbol{g}\left(\boldsymbol{x}^{k_l}\right)\right\}\right\| + \left\|\bar{\boldsymbol{v}}\right\|' \left\|\boldsymbol{h}\left(\boldsymbol{x}^{k_l}\right)\right\|
\end{aligned}$$

$$
\begin{aligned}
&+ \left\| \left(\bar{\boldsymbol{u}}_{\boldsymbol{2}}^T, \bar{\boldsymbol{v}}_{\boldsymbol{2}}^T \right)^T \right\|' \left\| \left(\max \left\{ \boldsymbol{0}, \boldsymbol{g_2} \left(\boldsymbol{x}^{k_l} \right) \right\}^T, \boldsymbol{h_2} \left(\boldsymbol{x}^{k_l} \right)^T \right)^T \right\| \\
&\geq f \left(\boldsymbol{x}^{k_l} \right) + \bar{\boldsymbol{u}}^T \max \left\{ \boldsymbol{0}, \boldsymbol{g} \left(\boldsymbol{x}^{k_l} \right) \right\} + \bar{\boldsymbol{v}}^T \boldsymbol{h} \left(\boldsymbol{x}^{k_l} \right) \\
&+ \bar{\boldsymbol{u}}_{\boldsymbol{2}}^T \max \left\{ \boldsymbol{0}, \boldsymbol{g_2} \left(\boldsymbol{x}^{k_l} \right) \right\} + \bar{\boldsymbol{v}}_{\boldsymbol{2}}^T \boldsymbol{h_2} \left(\boldsymbol{x}^{k_l} \right).
\end{aligned}
$$

Since $\boldsymbol{g}(\boldsymbol{x}) \leq \max\{\boldsymbol{0}, \boldsymbol{g}(\boldsymbol{x})\}$ and $\boldsymbol{g_2}(\boldsymbol{x}) \leq \max\{\boldsymbol{0}, \boldsymbol{g_2}(\boldsymbol{x})\}$, it follows that

$$
\begin{aligned}
F_{AE} \left(\boldsymbol{x}^{k_l}; \mu, \mu_2 \right) &\geq f \left(\boldsymbol{x}^{k_l} \right) + \bar{\boldsymbol{u}}^T \boldsymbol{g} \left(\boldsymbol{x}^{k_l} \right) + \bar{\boldsymbol{v}}^T \boldsymbol{h} \left(\boldsymbol{x}^{k_l} \right) \\
&+ \bar{\boldsymbol{u}}_{\boldsymbol{2}}^T \boldsymbol{g_2} \left(\boldsymbol{x}^{k_l} \right) + \bar{\boldsymbol{v}}_{\boldsymbol{2}}^T \boldsymbol{h_2} \left(\boldsymbol{x}^{k_l} \right) \\
&= \mathcal{L} \left(\boldsymbol{x}^{k_l}; \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}, \bar{\boldsymbol{u}}_{\boldsymbol{2}}, \bar{\boldsymbol{v}}_{\boldsymbol{2}} \right).
\end{aligned}
$$

By using (1.33), for sufficiently large k_l, we get

$$
\begin{aligned}
F_{AE} \left(\boldsymbol{x}^{k_l}; \mu, \mu_2 \right) &\geq \mathcal{L} \left(\boldsymbol{x}^{k_l}; \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}, \bar{\boldsymbol{u}}_{\boldsymbol{2}}, \bar{\boldsymbol{v}}_{\boldsymbol{2}} \right) \\
&> \mathcal{L} \left(\bar{\boldsymbol{x}}; \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}, \bar{\boldsymbol{u}}_{\boldsymbol{2}}, \bar{\boldsymbol{v}}_{\boldsymbol{2}} \right) \\
&= f(\bar{\boldsymbol{x}}) \\
&= F_{AE}(\bar{\boldsymbol{x}}; \mu, \mu_2) \\
&\geq F_{AE} \left(\boldsymbol{x}^{k_l}; \mu, \mu_2 \right),
\end{aligned}
$$

which is a contradiction. Thus, $\bar{\boldsymbol{x}}$ is a strict local solution to (AEP). This means that for $\mu > \bar{\mu}$, and $\mu_2 > \bar{\mu}_2$ there exists an $\varepsilon > 0$ such that for all $\boldsymbol{x} \in \{\boldsymbol{x} \in \mathbb{R}^n : \ \|\boldsymbol{x} - \bar{\boldsymbol{x}}\| \leq \varepsilon\} =: \mathcal{B}(\bar{\boldsymbol{x}}, \varepsilon)$

$$
F_{AE}(\bar{\boldsymbol{x}}; \mu, \mu_2) < F_{AE}(\boldsymbol{x}; \mu, \mu_2),
$$

and also for all $\boldsymbol{x} \in \mathcal{B}(\bar{\boldsymbol{x}}, \varepsilon) \cap \mathcal{X}$. Since $\left\| \left(\max\{\boldsymbol{0}, \boldsymbol{g_2}(\boldsymbol{x})\}^T, \boldsymbol{h_2}(\boldsymbol{x})^T \right)^T \right\| = 0$ for all $\boldsymbol{x} \in \mathcal{B}(\bar{\boldsymbol{x}}, \varepsilon) \cap \mathcal{X}$,

$$
F_E(\bar{\boldsymbol{x}}; \mu) < F_E(\boldsymbol{x}; \mu).
$$

This completes the proof. □

For further purposes, consider the following reformulation of the original problem (OP) with multiple inequality constraints

$$
\begin{aligned}
\min \quad & f(\boldsymbol{x}) \\
\text{subject to} \quad & \boldsymbol{g}_i(\boldsymbol{x}) \leq \boldsymbol{0} \quad i = 1, \ldots, k \\
& \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0} \\
& \boldsymbol{x} \in \mathcal{X},
\end{aligned}
\tag{MCOP}
$$

where $\boldsymbol{g}_i(\boldsymbol{x}) = (g_{i1}(\boldsymbol{x}), \ldots, g_{im_i}(\boldsymbol{x}))^T$, $i = 1, \ldots, k$. The corresponding objective function to the exact penalty function problem (EP) is as follows

$$
F_E(\boldsymbol{x}; \mu) = f(\boldsymbol{x}) + \sum_{i=1}^{k} \mu_i \|\max\{\boldsymbol{0}, \boldsymbol{g}_i(\boldsymbol{x})\}\| + \mu \|\boldsymbol{h}(\boldsymbol{x})\|. \tag{1.34}
$$

Furthermore, denote

$$
I_i(\bar{\boldsymbol{x}}) = \{j \in (1, \ldots, m_i) : g_{ij}(\bar{\boldsymbol{x}}) = 0\}, \ i = 1, \ldots, k.
$$

In a special case of $f(\boldsymbol{x})$, $g_{ij}(\boldsymbol{x})$, $j = 1, \ldots, m_i$, $i = 1, \ldots, k$, and $h_i(\boldsymbol{x})$, $i = 1, \ldots, l$, and $\mathcal{X}$, any solution to problem with multiple constraints (MCOP) is a global solution to the exact penalty function problem (EP) with the objective function defined by (1.34) for sufficiently large μ, and vice versa.

Theorem 1.3. *Suppose that*

$$\mathcal{X} = \{\boldsymbol{x} \colon g_i(\boldsymbol{x}) \leq 0,\ i = 1, \dots, M,\ h_i(\boldsymbol{x}) = 0,\ i = l+1, \dots, L\},$$

is a nonempty set, $\mathcal{C} \supset \mathcal{X}$ is an open convex set, and $\bar{\boldsymbol{x}} \in S$ is a KKT point for problem (MCOP) *with Lagrangian multipliers $\bar{u}_{ij}$, $j = 1, \dots, m_i$, $i = 1, \dots, k$, and $\bar{v}_i$, $i = 1, \dots, l$ associated with the inequality $g_{ij}(\boldsymbol{x})$, $j = 1, \dots, m_i$, $i = 1, \dots, k$, and equality constraints $h_i(\boldsymbol{x})$, $i = 1, \dots, l$, respectively.*

Denote

(i) $\bar{\boldsymbol{u}}_i = (\bar{u}_{i1}, \dots, \bar{u}_{im_i})^T$, $i = 1, \dots, k$, $\bar{\boldsymbol{v}} = (\bar{v}_1, \dots, \bar{v}_l)^T$,

(ii) $\bar{u}_i$, $i = 1, \dots, M$, *and* $\bar{v}_i$, $i = l+1, \dots, L$ *the Lagrangian multipliers associated with* $g_i(\boldsymbol{x})$, $i = 1, \dots, M$, *and* $h_i(\boldsymbol{x})$, $i = l+1, \dots, L$, *respectively,*

(iii) $I_{\mathcal{X}}(\bar{\boldsymbol{x}}) = \{i \in (1, \dots, M) : g_i(\bar{\boldsymbol{x}}) = 0\}$,

(iv) $I^+(\bar{\boldsymbol{x}}) = \{i \in (1, \dots, l) : \bar{v}_i > 0\}$, $I^+_{\mathcal{X}}(\bar{\boldsymbol{x}}) = \{i \in (l+1, \dots, L) : \bar{v}_i > 0\}$,

(v) $I^-(\bar{\boldsymbol{x}}) = \{i \in (1, \dots, l) : \bar{v}_i < 0\}$, $I^-_{\mathcal{X}}(\bar{\boldsymbol{x}}) = \{i \in (l+1, \dots, L) : \bar{v}_i < 0\}$.

Furthermore, suppose that $f(\boldsymbol{x})$ is a convex function on $\mathcal{C}$,

(i) $g_{ij}(\boldsymbol{x})$, $j \in I_i(\bar{\boldsymbol{x}})$, $i = 1, \dots, k$, $g_i(\boldsymbol{x})$, $i \in I_{\mathcal{X}}(\bar{\boldsymbol{x}})$ *are convex functions on* $\mathcal{C}$,

(ii) h_i, $i \in I^+(\bar{\boldsymbol{x}}) \cup I^+_{\mathcal{X}}(\bar{\boldsymbol{x}})$ *are convex functions on* $\mathcal{C}$,

(iii) h_i, $i \in I^-(\bar{\boldsymbol{x}}) \cup I^-_{\mathcal{X}}(\bar{\boldsymbol{x}})$ *are concave functions on* $\mathcal{C}$.

Then, for

$$\mu_i \geq \bar{\mu}_i = \|\bar{\boldsymbol{u}}_i\|',\ i = 1, \dots, k \text{ and } \mu \geq \bar{\mu} = \|\bar{\boldsymbol{v}}\|'$$

where $\|\cdot\|'$ is the dual norm to $\|\cdot\|$, $\bar{\boldsymbol{x}}$ also minimises the exact penalty objective function (1.34)*.*

Moreover, if $\bar{\boldsymbol{x}}_E$ is solution to the exact penalty problem (EP) *with objective function* (1.34) *for $\mu_i > \bar{\mu}_i = \|\bar{\boldsymbol{u}}_i\|'$, $i = 1, \dots, k$ and $\mu > \|\bar{\boldsymbol{v}}\|'$, then $\bar{\boldsymbol{x}}_E$ is also a solution to the problem with multiple inequality constraints* (MCOP)*.*

Proof. Denote $\boldsymbol{g}_{\mathcal{X}}(\boldsymbol{x}) = (g_1(\boldsymbol{x}), \dots, g_M(\boldsymbol{x}))^T$, $\boldsymbol{h}_{\mathcal{X}}(\boldsymbol{x}) = (h_{l+1}(\boldsymbol{x}), \dots, h_L(\boldsymbol{x}))^T$, $\bar{\boldsymbol{u}}_{\mathcal{X}} = (\bar{u}_1^2, \dots, \bar{u}_M^2)^T$, and $\bar{\boldsymbol{v}}_{\mathcal{X}} = (\bar{v}_{l+1}^2, \dots, \bar{v}_L^2)^T$.

Due to our assumptions, $\mathcal{L}(\boldsymbol{x}, \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}})$ is convex on $\mathcal{C}$ and differentiable at $\bar{\boldsymbol{x}}$. This implies that

$$\mathcal{L}(\boldsymbol{x}, \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}) - \mathcal{L}(\bar{\boldsymbol{x}}, \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}) \geq \nabla_{\boldsymbol{x}} \mathcal{L}(\bar{\boldsymbol{x}}, \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}})^T (\boldsymbol{x} - \bar{\boldsymbol{x}}),$$

vide Theorem 3.2.5 and Lemma 3.3.2 in [3]. Since $\bar{\boldsymbol{x}}$ is a KKT point, $\nabla_{\boldsymbol{x}} \mathcal{L}(\bar{\boldsymbol{x}}, \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}) = \boldsymbol{0}$. Thus, we can write

$$\mathcal{L}(\boldsymbol{x}, \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}) \geq \mathcal{L}(\bar{\boldsymbol{x}}, \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}). \tag{1.35}$$

By using the generalised Cauchy–Schwarz inequality (A.1), we obtain

$$\begin{aligned} F_E(\boldsymbol{x};\mu) &= f(\boldsymbol{x}) + \sum_{i=1}^{k} \mu_i \left\|\max\{\boldsymbol{0}, \boldsymbol{g}_i(\boldsymbol{x})\}\right\| + \mu \left\|\boldsymbol{h}(\boldsymbol{x})\right\| \\ &\geq f(\boldsymbol{x}) + \sum_{i=1}^{k} \left\|\bar{\boldsymbol{u}}_i\right\|' \left\|\max\{\boldsymbol{0}, \boldsymbol{g}_i(\boldsymbol{x})\}\right\| + \left\|\bar{\boldsymbol{v}}\right\|' \left\|\boldsymbol{h}(\boldsymbol{x})\right\| \\ &\geq f(\boldsymbol{x}) + \sum_{i=1}^{k} \bar{\boldsymbol{u}}_i^T \max\{\boldsymbol{0}, \boldsymbol{g}_i(\boldsymbol{x})\} + \bar{\boldsymbol{v}}^T \boldsymbol{h}(\boldsymbol{x}). \end{aligned}$$

Since $\boldsymbol{g}_i(\boldsymbol{x}) \leq \max\{\boldsymbol{0}, \boldsymbol{g}_i(\boldsymbol{x})\}$, $i = 1, \ldots, k$, it follows that

$$F_E(\boldsymbol{x};\mu) \geq f(\boldsymbol{x}) + \sum_{i=1}^{k} \bar{\boldsymbol{u}}_i^T \boldsymbol{g}_i(\boldsymbol{x}) + \bar{\boldsymbol{v}}^T \boldsymbol{h}(\boldsymbol{x}).$$

Since $\boldsymbol{g}_{\mathcal{X}}(\boldsymbol{x}) \leq \boldsymbol{0}$ and $\boldsymbol{h}_{\mathcal{X}}(\boldsymbol{x}) = \boldsymbol{0}$ for all $\boldsymbol{x} \in \mathcal{X}$, we can write

$$\begin{aligned} f(\boldsymbol{x}) + \sum_{i=1}^{k} \bar{\boldsymbol{u}}_i^T \boldsymbol{g}_i(\boldsymbol{x}) + \bar{\boldsymbol{v}}^T \boldsymbol{h}(\boldsymbol{x}) &\geq f(\boldsymbol{x}) + \sum_{i=1}^{k} \bar{\boldsymbol{u}}_i^T \boldsymbol{g}_i(\boldsymbol{x}) + \bar{\boldsymbol{v}}^T \boldsymbol{h}(\boldsymbol{x}) \\ &+ \bar{\boldsymbol{u}}_{\mathcal{X}}^T \boldsymbol{g}_{\mathcal{X}}(\boldsymbol{x}) + \bar{\boldsymbol{v}}_{\mathcal{X}}^T \boldsymbol{h}_{\mathcal{X}}(\boldsymbol{x}) = \mathcal{L}(\boldsymbol{x}, \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}) \end{aligned}$$

By using (1.35), for all $\boldsymbol{x} \in \mathcal{X}$ we get

$$\begin{aligned} F_E(\boldsymbol{x};\mu) &\geq \mathcal{L}(\boldsymbol{x}, \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}) \\ &\geq \mathcal{L}(\bar{\boldsymbol{x}}, \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}) \\ &= f(\bar{\boldsymbol{x}}) \\ &= F_E(\bar{\boldsymbol{x}};\mu), \end{aligned}$$

thus $\bar{\boldsymbol{x}}$ is a solution to the exact penalty problem (EP) for $\mu_i \geq \bar{\mu}_i$, $i = 1, \ldots, k$ and $\mu \geq \bar{\mu}$.

Since $\bar{\boldsymbol{x}}_E$ solves (EP), we have

$$f(\bar{\boldsymbol{x}}_E) + \sum_{i=1}^{k} \mu_i \left\|\max\{\boldsymbol{0}, \boldsymbol{g}_i(\bar{\boldsymbol{x}}_E)\}\right\| + \mu \left\|\boldsymbol{h}(\bar{\boldsymbol{x}}_E)\right\| = F_E(\bar{\boldsymbol{x}}_E;\mu) \leq F_E(\bar{\boldsymbol{x}};\mu) = f(\bar{\boldsymbol{x}}),$$

and so

$$f(\bar{\boldsymbol{x}}_E) \leq f(\bar{\boldsymbol{x}}) - \sum_{i=1}^{k} \mu_i \left\|\max\{\boldsymbol{0}, \boldsymbol{g}_i(\bar{\boldsymbol{x}}_E)\}\right\| - \mu \left\|\boldsymbol{h}(\bar{\boldsymbol{x}}_E)\right\|. \tag{1.36}$$

Since $f(\boldsymbol{x})$, $g_{ij}(\boldsymbol{x})$, $j \in I_i(\bar{\boldsymbol{x}})$, $i = 1, \ldots, k$, $g_i(\boldsymbol{x})$, $i \in I_{\mathcal{X}}(\bar{\boldsymbol{x}})$, and $h_i(\boldsymbol{x})$, $i \in I^+(\bar{\boldsymbol{x}}) \cup I_{\mathcal{X}}^+(\bar{\boldsymbol{x}})$ are convex functions, it holds true for all $\boldsymbol{x} \in \mathcal{X}$

$$f(\boldsymbol{x}) \geq f(\bar{\boldsymbol{x}}) + \nabla_{\boldsymbol{x}} f(\bar{\boldsymbol{x}})^T (\boldsymbol{x} - \bar{\boldsymbol{x}}), \tag{1.37}$$

$$g_{ij}(\boldsymbol{x}) \geq g_{ij}(\bar{\boldsymbol{x}}) + \nabla_{\boldsymbol{x}} g_{ij}(\bar{\boldsymbol{x}})^T (\boldsymbol{x} - \bar{\boldsymbol{x}}) \quad \forall\, j \in I_i(\bar{\boldsymbol{x}}),\ i = 1, \ldots, k, \tag{1.38}$$

$$g_i(\boldsymbol{x}) \geq g_i(\bar{\boldsymbol{x}}) + \nabla_{\boldsymbol{x}} g_i(\bar{\boldsymbol{x}})^T (\boldsymbol{x} - \bar{\boldsymbol{x}}) \quad \forall\, i \in I_{\mathcal{X}}(\bar{\boldsymbol{x}}), \tag{1.39}$$

$$h_i(\boldsymbol{x}) \geq h_i(\bar{\boldsymbol{x}}) + \nabla_{\boldsymbol{x}} h_i(\bar{\boldsymbol{x}})^T (\boldsymbol{x} - \bar{\boldsymbol{x}}) \quad \forall\, i \in I^+(\bar{\boldsymbol{x}}) \cup I_{\mathcal{X}}^+(\bar{\boldsymbol{x}}). \tag{1.40}$$

Besides, $h_i(\boldsymbol{x})$, $i \in I^-(\bar{\boldsymbol{x}}) \cup I^-_{\mathcal{X}}(\bar{\boldsymbol{x}})$ are concave, and hence, the following holds true for all $\boldsymbol{x} \in \mathcal{X}$

$$h_i(\boldsymbol{x}) \leq h_i(\bar{\boldsymbol{x}}) + \nabla_{\boldsymbol{x}} h_i(\bar{\boldsymbol{x}})^T (\boldsymbol{x} - \bar{\boldsymbol{x}}) \quad \forall\, i \in I^-(\bar{\boldsymbol{x}}) \cup I^-_{\mathcal{X}}(\bar{\boldsymbol{x}}). \tag{1.41}$$

Suppose that $\bar{\boldsymbol{x}}_E$ is not feasible for (OP). Using (1.37), we have

$$f(\bar{\boldsymbol{x}}_E) \geq f(\bar{\boldsymbol{x}}) + \nabla_{\boldsymbol{x}} f(\bar{\boldsymbol{x}})^T (\bar{\boldsymbol{x}}_E - \bar{\boldsymbol{x}}). \tag{1.42}$$

Denote

$$\begin{gathered} J_{\boldsymbol{g}_i}(\bar{\boldsymbol{x}}) = \left(\nabla_{\boldsymbol{x}} g_{i1}(\bar{\boldsymbol{x}}), \ldots, \nabla_{\boldsymbol{x}} g_{im_i}(\bar{\boldsymbol{x}})\right)^T,\ i = 1, \ldots, k, \\ J_{\boldsymbol{h}}(\bar{\boldsymbol{x}}) = \left(\nabla_{\boldsymbol{x}} h_1(\bar{\boldsymbol{x}}), \ldots, \nabla_{\boldsymbol{x}} h_l(\bar{\boldsymbol{x}})\right)^T \end{gathered}$$

and

$$\begin{gathered} J_{\boldsymbol{g}_{\mathcal{X}}}(\bar{\boldsymbol{x}}) = \left(\nabla_{\boldsymbol{x}} g_1(\bar{\boldsymbol{x}}), \ldots, \nabla_{\boldsymbol{x}} g_M(\bar{\boldsymbol{x}})\right)^T, \\ J_{h_{\mathcal{X}}}(\bar{\boldsymbol{x}}) = \left(\nabla_{\boldsymbol{x}} h_{l+1}(\bar{\boldsymbol{x}}), \ldots, \nabla_{\boldsymbol{x}} h_L(\bar{\boldsymbol{x}})\right)^T. \end{gathered}$$

Since $\bar{\boldsymbol{x}}$ is a KKT point, we can write

$$\nabla_{\boldsymbol{x}} f(\bar{\boldsymbol{x}})^T = -\sum_{i=1}^{K} \bar{\boldsymbol{u}}_i^T J_{\boldsymbol{g}_i}(\bar{\boldsymbol{x}}) - \bar{\boldsymbol{v}}^T J_{\boldsymbol{h}}(\bar{\boldsymbol{x}}) - \bar{\boldsymbol{u}}_{\mathcal{X}}^T J_{\boldsymbol{g}_{\mathcal{X}}}(\bar{\boldsymbol{x}}) - \bar{\boldsymbol{v}}_{\mathcal{X}}^T J_{\boldsymbol{h}_{\mathcal{X}}}(\bar{\boldsymbol{x}}). \tag{1.43}$$

By using (1.43) in (1.42), we obtain

$$\begin{aligned} f(\bar{\boldsymbol{x}}_E) \geq f(\bar{\boldsymbol{x}}) &- \sum_{i=1}^{k} \bar{\boldsymbol{u}}_i^T J_{\boldsymbol{g}_i}(\bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_E - \bar{\boldsymbol{x}}) - \bar{\boldsymbol{v}}^T J_{\boldsymbol{h}}(\bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_E - \bar{\boldsymbol{x}}) \\ &- \bar{\boldsymbol{u}}_{\mathcal{X}}^T J_{\boldsymbol{g}_{\mathcal{X}}}(\bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_E - \bar{\boldsymbol{x}}) - \bar{\boldsymbol{v}}_{\mathcal{X}}^T J_{\boldsymbol{h}_{\mathcal{X}}}(\bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_E - \bar{\boldsymbol{x}}). \end{aligned}$$

By using (1.38), (1.39), (1.40) and (1.41), we get

$$\begin{aligned} f(\bar{\boldsymbol{x}}_E) \geq f(\bar{\boldsymbol{x}}) &+ \sum_{i=1}^{k} \bar{\boldsymbol{u}}_i^T (\boldsymbol{g}_i(\bar{\boldsymbol{x}}) - \boldsymbol{g}_i(\bar{\boldsymbol{x}}_E)) + \bar{\boldsymbol{v}}^T (\boldsymbol{h}(\bar{\boldsymbol{x}}) - \boldsymbol{h}(\bar{\boldsymbol{x}}_E)) \\ &+ \bar{\boldsymbol{u}}_{\mathcal{X}}^T (\boldsymbol{g}_{\mathcal{X}}(\bar{\boldsymbol{x}}) - \boldsymbol{g}_{\mathcal{X}}(\bar{\boldsymbol{x}}_E)) + \bar{\boldsymbol{v}}_{\mathcal{X}}^T (\boldsymbol{h}_{\mathcal{X}}(\bar{\boldsymbol{x}}) - \boldsymbol{h}_{\mathcal{X}}(\bar{\boldsymbol{x}}_E)). \end{aligned}$$

Since $\bar{\boldsymbol{u}}_i^T \boldsymbol{g}_i(\bar{\boldsymbol{x}}) = 0$, $i = 1, \ldots, k$, $\bar{\boldsymbol{v}}^T \boldsymbol{h}(\bar{\boldsymbol{x}}) = 0$, $\bar{\boldsymbol{u}}_{\mathcal{X}}^T \boldsymbol{g}_{\mathcal{X}}(\bar{\boldsymbol{x}}) = 0$ and $\bar{\boldsymbol{v}}_{\mathcal{X}}^T \boldsymbol{h}_{\mathcal{X}}(\bar{\boldsymbol{x}}) = 0$, and $\boldsymbol{g}_{\mathcal{X}}(\bar{\boldsymbol{x}}_E) \leq \mathbf{0}$, $\boldsymbol{h}_{\mathcal{X}}(\bar{\boldsymbol{x}}_E) = \mathbf{0}$, we can write

$$\begin{aligned} f(\bar{\boldsymbol{x}}) &+ \sum_{i=1}^{k} \bar{\boldsymbol{u}}_i^T (\boldsymbol{g}_i(\bar{\boldsymbol{x}}) - \boldsymbol{g}_i(\bar{\boldsymbol{x}}_E)) + \bar{\boldsymbol{v}}^T (\boldsymbol{h}(\bar{\boldsymbol{x}}) - \boldsymbol{h}(\bar{\boldsymbol{x}}_E)) \\ &+ \bar{\boldsymbol{u}}_{\mathcal{X}}^T (\boldsymbol{g}_{\mathcal{X}}(\bar{\boldsymbol{x}}) - \boldsymbol{g}_{\mathcal{X}}(\bar{\boldsymbol{x}}_E)) + \bar{\boldsymbol{v}}_{\mathcal{X}}^T (\boldsymbol{h}_{\mathcal{X}}(\bar{\boldsymbol{x}}) - \boldsymbol{h}_{\mathcal{X}}(\bar{\boldsymbol{x}}_E)) \\ &\geq f(\bar{\boldsymbol{x}}) - \sum_{i=1}^{k} \bar{\boldsymbol{u}}_i^T \boldsymbol{g}_i(\bar{\boldsymbol{x}}_E) - \bar{\boldsymbol{v}}^T \boldsymbol{h}(\bar{\boldsymbol{x}}_E). \end{aligned}$$

Since $\boldsymbol{g}_i(\boldsymbol{x}) \leq \max\{\mathbf{0}, \boldsymbol{g}_i(\boldsymbol{x})\}$, $i = 1, \ldots, k$, it follows that

$$f(\bar{\boldsymbol{x}}) - \sum_{i=1}^{k} \bar{\boldsymbol{u}}_i^T \boldsymbol{g}_i(\bar{\boldsymbol{x}}_E) - \bar{\boldsymbol{v}}^T \boldsymbol{h}(\bar{\boldsymbol{x}}_E) \geq f(\bar{\boldsymbol{x}}) - \sum_{i=1}^{k} \bar{\boldsymbol{u}}_i^T \max\{\mathbf{0}, \boldsymbol{g}_i(\bar{\boldsymbol{x}}_E)\} - \bar{\boldsymbol{v}}^T \boldsymbol{h}(\bar{\boldsymbol{x}}_E).$$

After using the generalised Cauchy–Schwarz inequality (A.1), we can write

$$\begin{aligned} f(\bar{\boldsymbol{x}}_E) &\geq f(\bar{\boldsymbol{x}}) - \sum_{i=1}^{k} \bar{\boldsymbol{u}}_i^T \max\{\boldsymbol{0}, \boldsymbol{g}_i(\bar{\boldsymbol{x}}_E)\} - \bar{\boldsymbol{v}}^T \boldsymbol{h}(\bar{\boldsymbol{x}}_E) \\ &\geq f(\bar{\boldsymbol{x}}) - \sum_{i=1}^{k} \|\bar{\boldsymbol{u}}_i\|' \, \|\max\{\boldsymbol{0}, \boldsymbol{g}_i(\bar{\boldsymbol{x}}_E)\}\| - \|\bar{\boldsymbol{v}}\|' \, \|\boldsymbol{h}(\bar{\boldsymbol{x}}_E)\| \\ &> f(\bar{\boldsymbol{x}}) - \sum_{i=1}^{k} \mu_i \, \|\max\{\boldsymbol{0}, \boldsymbol{g}_i(\bar{\boldsymbol{x}}_E)\}\| - \mu \, \|\boldsymbol{h}(\bar{\boldsymbol{x}}_E)\|, \end{aligned}$$

which contradicts (1.36).Therefore, $\bar{\boldsymbol{x}}_E$ is feasible for (MCOP), and

$$f(\bar{\boldsymbol{x}}_E) \leq f(\bar{\boldsymbol{x}}).$$

Since $\bar{\boldsymbol{x}}$ is a KKT point for (MCOP), it is an optimal solution to (MCOP), vide [3, Theorem 4.3.8], and therefore $\bar{\boldsymbol{x}}_E$ is also and optimal solution to (MCOP). This completes the proof. □

Note that Theorem 1.3 also works with the original problem (OP) and the corresponding exact penalty function problem (EP) with a common penalty parameter μ. In such a case, the threshold $\bar{\mu}$ equals to

$$\bar{\mu} = \max\left\{\|\bar{\boldsymbol{u}}\|', \|\bar{\boldsymbol{v}}\|'\right\}.$$

Evidently, the assumption of Theorem 1.3 must be satisfied, i.e. the convexity and concavity requirements on $f(\boldsymbol{x})$, $g_i(\boldsymbol{x})$, $h_i(\boldsymbol{x})$ and $\mathcal{X}$ must be given as in the theorem.

1.3.1 Invex functions

Generalised convex functions have been introduced in order to lessen as much as possible the convexity requirements for results related to optimisation problems. This motivation led to put forward and employ pseudo-convex, e.g. [23, 24], and quasi-convex function, e.g. [22].

Invex functions intend to generalise convex functions as well. Dealing with KKT conditions and Wolfe duality, vide [36], Hanson noted that the usual convexity requirements can be altered. By substituting the linear term $(\boldsymbol{x} - \boldsymbol{y})$ appearing in the definition of differentiable convex functions by an arbitrary vector-valued function $\eta(\boldsymbol{x}, \boldsymbol{y})$, he introduced a new class of differentiable functions, vide [18].

Definiton 1.2 (Invex function)**.** Assume $\mathcal{O} \subseteq \mathbb{R}^n$ is an open set. A differentiable function $f(\boldsymbol{x}) : \mathcal{O} \to \mathbb{R}$ is invex on $\mathcal{O}$ with respect to $\eta(\boldsymbol{x}, \boldsymbol{y})$ if there exists a vector-valued function $\eta(\boldsymbol{x}, \boldsymbol{y}) : \mathcal{O} \times \mathcal{O} \to \mathbb{R}^n$ such that

$$f(\boldsymbol{x}) - f(\boldsymbol{y}) \geq \eta(\boldsymbol{x}, \boldsymbol{y})^T \nabla_{\boldsymbol{y}} f(\boldsymbol{y}), \quad \forall\, \boldsymbol{x}, \boldsymbol{y} \in \mathcal{O}. \tag{1.44}$$

The vector-valued function $\eta(\boldsymbol{x}, \boldsymbol{y})$ is sometimes referred to as the *kernel function*. The term *invex* is due to Craven [11] and is an abbreviation of *invariant convex*, since one can construct an invex function with respect to $\eta(\boldsymbol{x}, \boldsymbol{y})$ by the following way, cf. [25]:

Let $g(\boldsymbol{x}) : \mathbb{R}^n \to \mathbb{R}$ be differentiable convex and $\boldsymbol{\Phi}(\boldsymbol{y}) : \mathbb{R}^r \to \mathbb{R}^n$, $r \geq n$, be differentiable with

$$J_{\boldsymbol{\Phi}}(\boldsymbol{y}) = \left(\nabla_{\boldsymbol{y}}\Phi_1(\boldsymbol{y}), \ldots, \nabla_{\boldsymbol{y}}\Phi_n(\boldsymbol{y})\right)^T$$

of rank n. Then $f(\boldsymbol{y}) = g(\boldsymbol{\Phi}(\boldsymbol{y}))$ is invex since $\forall\ \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^r$, we have

$$f(\boldsymbol{x}) - f(\boldsymbol{y}) = g(\boldsymbol{\Phi}(\boldsymbol{x})) - g(\boldsymbol{\Phi}(\boldsymbol{y})) \geq (\boldsymbol{\Phi}(\boldsymbol{x}) - \boldsymbol{\Phi}(\boldsymbol{y}))^T \nabla_{\boldsymbol{x}} g(\boldsymbol{\Phi}(\boldsymbol{y})).$$

As $J_\Phi(\boldsymbol{y})$ is full of row rank, equation

$$(\boldsymbol{\Phi}(\boldsymbol{x}) - \boldsymbol{\Phi}(\boldsymbol{y})) = J_{\boldsymbol{\Phi}}(\boldsymbol{y})\,\eta(\boldsymbol{x}, \boldsymbol{y}) \tag{1.45}$$

has a solution $\eta(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^r$. By multiplying equation (1.45) by $\nabla_{\boldsymbol{x}} g(\boldsymbol{\Phi}(\boldsymbol{y}))^T$, we get

$$\nabla_{\boldsymbol{x}} g(\boldsymbol{\Phi}(\boldsymbol{y}))^T (\boldsymbol{\Phi}(\boldsymbol{x}) - \boldsymbol{\Phi}(\boldsymbol{y})) = \nabla_{\boldsymbol{x}} g(\boldsymbol{\Phi}(\boldsymbol{y}))^T J_{\boldsymbol{\Phi}}(\boldsymbol{y})\,\eta(\boldsymbol{x}, \boldsymbol{y}). \tag{1.46}$$

Since $\nabla_{\boldsymbol{y}} f(\boldsymbol{y}) = J_{\boldsymbol{\Phi}}(\boldsymbol{y})^T \nabla_{\boldsymbol{x}} g(\boldsymbol{\Phi}(\boldsymbol{y}))$, equation (1.46) is equivalent to

$$(\boldsymbol{\Phi}(\boldsymbol{x}) - \boldsymbol{\Phi}(\boldsymbol{y}))^T \nabla_{\boldsymbol{x}} g(\boldsymbol{\Phi}(\boldsymbol{y})) = \eta(\boldsymbol{x}, \boldsymbol{y})^T \nabla_{\boldsymbol{y}} f(\boldsymbol{y})$$

Hence,

$$f(\boldsymbol{x}) - f(\boldsymbol{y}) \geq \eta(\boldsymbol{x}, \boldsymbol{y})^T \nabla_{\boldsymbol{y}} f(\boldsymbol{y}), \quad \forall\ \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^r$$

and for some $\eta(\boldsymbol{x}, \boldsymbol{y}) : \mathbb{R}^r \times \mathbb{R}^r \to \mathbb{R}^r$.

Ben-Israel and Mond [4] showed that a differentiable function is invex with respect to $\eta(\boldsymbol{x}, \boldsymbol{y})$ if and only if every stationary point is a global minimum. Hence, invex functions with respect to the same $\eta(\boldsymbol{x}, \boldsymbol{y})$ were constructed in order for the KKT condition to be sufficient for a global minimum, cf. [25].

Obviously, the particular case of differentiable convex function is obtained from (1.44) by putting $\eta(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{x} - \boldsymbol{y}$. Thus, invexity is commonly considered as a generalisation of convexity. However, this notion can lead to misapprehensions as it is pointless to discuss invexity without considering the kernel function.

The following illustrative example of a invex function which is not convex is taken from [21]. Consider the function $f(x) = x + \sin(x)$ defined on $\left(0, \frac{\pi}{2}\right)$. The $f(x)$ is invex with respect to

$$\eta(x, y) = \frac{\sin(x) - \sin(y)}{\cos(y)},$$

but is not convex as for $x = \frac{\pi}{4}$ and $y = \frac{\pi}{6}$

$$f(x) - f(y) \ngeq (x - y)\nabla_y f(y).$$

An example of an invex program which is not convex can be found in [18].

Preparatory to the next subsection, let us define the negative of an invex function with respect to $\eta(\boldsymbol{x}, \boldsymbol{y})$, the incave function with respect to $\eta(\boldsymbol{x}, \boldsymbol{y})$.

Definiton 1.3 (Incave function)**.** Assume $\mathcal{O} \subseteq \mathbb{R}^n$ is an open set. A differentiable function $f(\boldsymbol{x}) : \mathcal{O} \to \mathbb{R}$ is incave on $\mathcal{O}$ with respect to $\eta(\boldsymbol{x}, \boldsymbol{y})$ if $-f(\boldsymbol{x})$ is invex on $\mathcal{O}$ with respect to $\eta(\boldsymbol{x}, \boldsymbol{y})$.

1.3.2 Invex functions and exact penalty function methods

It is of utmost importance to find ways for employing exact penalty functions for non-convex problems. Antczak [1] recently proposed to use invex and incave functions with respect to the same $\eta(\boldsymbol{x},\boldsymbol{y})$ instead of convex and concave functions, respectively, for exact penalty methods. He established for the L_∞ penalty function that one can achieve similar results as in Theorem 1.3 by using invex and incave functions with respect to the same $\eta(\boldsymbol{x},\boldsymbol{y})$.

Let us reformulate Theorem 1.3 for invex and incave functions with respect to the same $\eta(\boldsymbol{x},\boldsymbol{y})$ and prove that the result likewise holds.

Theorem 1.4. *Suppose that*

$$\mathcal{X}=\{\boldsymbol{x}\colon g_i(\boldsymbol{x})\le 0,\ i=1,\dots,M,\ h_i(\boldsymbol{x})=0,\ i=l+1,\dots,L\},$$

is a nonempty set, $\mathcal{O}\supset\mathcal{X}$ is an open set, and $\bar{\boldsymbol{x}}\in S$ is a KKT point for problem (MCOP) *with Lagrangian multipliers $\bar{u}_{ij}$, $j=1,\dots,m_i$, $i=1,\dots,k$, and $\bar{v}_i$, $i=1,\dots,l$ associated with the inequality $g_{ij}(\boldsymbol{x})$, $j=1,\dots,m_i$, $i=1,\dots,k$, and equality constraints $h_i(\boldsymbol{x})$, $i=1,\dots,l$, respectively.*

Denote

(i) $\bar{\boldsymbol{u}}_i=(\bar{u}_{i1},\dots,\bar{u}_{im_i})^T$, $i=1,\dots,k$, $\bar{\boldsymbol{v}}=(\bar{v}_1,\dots,\bar{v}_l)^T$,

(ii) $\bar{u}_i$, $i=1,\dots,M$, *and* $\bar{v}_i$, $i=l+1,\dots,L$ *the Lagrangian multipliers associated with* $g_i(\boldsymbol{x})$, $i=1,\dots,M$, *and* $h_i(\boldsymbol{x})$, $i=l+1,\dots,L$, *respectively,*

(iii) $I_{\mathcal{X}}(\bar{\boldsymbol{x}})=\{i\in(1,\dots,M):g_i(\bar{\boldsymbol{x}})=0\}$,

(iv) $I^+(\bar{\boldsymbol{x}})=\{i\in(1,\dots,l):\bar{v}_i>0\}$, $I^+_{\mathcal{X}}(\bar{\boldsymbol{x}})=\{i\in(l+1,\dots,L):\bar{v}_i>0\}$,

(v) $I^-(\bar{\boldsymbol{x}})=\{i\in(1,\dots,l):\bar{v}_i<0\}$, $I^-_{\mathcal{X}}(\bar{\boldsymbol{x}})=\{i\in(l+1,\dots,L):\bar{v}_i<0\}$.

Furthermore, suppose that $f(\boldsymbol{x})$ is a invex functions on $\mathcal{O}$ with respect to a $\eta(\boldsymbol{x},\boldsymbol{y})$,

(i) $g_{ij}(\boldsymbol{x})$, $j\in I_i(\bar{\boldsymbol{x}})$, $i=1,\dots,k$, $g_i(\boldsymbol{x})$, $i\in I_{\mathcal{X}}(\bar{\boldsymbol{x}})$ *are invex functions on* $\mathcal{O}$ *with respect to the same* $\eta(\boldsymbol{x},\boldsymbol{y})$,

(ii) h_i, $i\in I^+(\bar{\boldsymbol{x}})\cup I^+_{\mathcal{X}}(\bar{\boldsymbol{x}})$ *are invex functions on* $\mathcal{O}$ *with respect to the same* $\eta(\boldsymbol{x},\boldsymbol{y})$,

(iii) h_i, $i\in I^-(\bar{\boldsymbol{x}})\cup I^-_{\mathcal{X}}(\bar{\boldsymbol{x}})$ *are incave on* $\mathcal{O}$ *functions with respect to the same* $\eta(\boldsymbol{x},\boldsymbol{y})$.

Then, for

$$\mu_i\ge\bar{\mu}_i=\|\bar{\boldsymbol{u}}_i\|',\ i=1,\dots,k\ \text{and}\ \mu\ge\|\bar{\boldsymbol{v}}\|'$$

where $\|\cdot\|'$ is the dual norm to $\|\cdot\|$, $\bar{\boldsymbol{x}}$ also minimises the exact penalty objective function (1.34)*.*

Moreover, if $\bar{\boldsymbol{x}}_E$ is solution to the exact penalty problem (EP) *with objective function* (1.34) *for $\mu_i>\bar{\mu}_i=\|\bar{\boldsymbol{u}}_i\|'$, $i=1,\dots,k$ and $\mu>\|\bar{\boldsymbol{v}}\|'$, then $\bar{\boldsymbol{x}}_E$ is also a solution to the problem with multiple inequality constraints* (MCOP)*.*

Proof. The proof of Theorem 1.4 is analogous to the proof of Theorem 1.3, and therefore we shall only discuss the validity of the main steps.

Due to our assumptions, $\mathcal{L}(\boldsymbol{x},\bar{\boldsymbol{u}},\bar{\boldsymbol{v}})$ is invex on $\mathcal{O}$ with respect to $\eta(\boldsymbol{x},\boldsymbol{y})$, cf. [25, Theorem 2.9]. This implies that

$$\mathcal{L}(\boldsymbol{x},\bar{\boldsymbol{u}},\bar{\boldsymbol{v}})-\mathcal{L}(\bar{\boldsymbol{x}},\bar{\boldsymbol{u}},\bar{\boldsymbol{v}})\ge\eta(\boldsymbol{x},\bar{\boldsymbol{x}})^T\nabla_{\boldsymbol{x}}\mathcal{L}(\bar{\boldsymbol{x}},\bar{\boldsymbol{u}},\bar{\boldsymbol{v}}).$$

Since $\bar{\boldsymbol{x}}$ is a KKT point, $\nabla_{\boldsymbol{x}}\mathcal{L}(\bar{\boldsymbol{x}},\bar{\boldsymbol{u}},\bar{\boldsymbol{v}}) = \mathbf{0}$. Thus, we can write

$$\mathcal{L}(\boldsymbol{x},\bar{\boldsymbol{u}},\bar{\boldsymbol{v}}) \geq \mathcal{L}(\bar{\boldsymbol{x}},\bar{\boldsymbol{u}},\bar{\boldsymbol{v}}).$$

Therefore, we have

$$F_E(\boldsymbol{x};\mu) \geq F_E(\bar{\boldsymbol{x}};\mu) \text{ for } \mu_i \geq \bar{\mu}_i,\ i=1,\dots,k,\ \mu \geq \bar{\mu}.$$

The second part of the theorem holds as well. Obviously, we get the same results by replacing $(\boldsymbol{x}-\bar{\boldsymbol{x}})$ with $\eta(\boldsymbol{x},\bar{\boldsymbol{x}})$. Namely, $\bar{\boldsymbol{x}}_E$ is feasible for (MCOP), and

$$f(\bar{\boldsymbol{x}}_E) \leq f(\bar{\boldsymbol{x}}).$$

Since $\bar{\boldsymbol{x}}$ is a KKT point for (MCOP), it is an optimal solution to (MCOP), cf. [18, Theorem 2.1], and therefore $\bar{\boldsymbol{x}}_E$ is also and optimal solution to (MCOP). This completes the proof. □

Note that for the existence of a common kernel function, the following system of inequalities

$$\begin{aligned}
f(\boldsymbol{x}) - f(\boldsymbol{y}) &\geq \eta(\boldsymbol{x},\boldsymbol{y})^T \nabla_{\boldsymbol{y}} f(\boldsymbol{y}),\\
g_{ij}(\boldsymbol{x}) - g_{ij}(\boldsymbol{y}) &\geq \eta(\boldsymbol{x},\boldsymbol{y})^T \nabla_{\boldsymbol{y}} g_{ij}(\boldsymbol{y}),\ j\in I_i(\bar{\boldsymbol{x}}),\ i=1,\dots,k,\\
g_i(\boldsymbol{x}) - g_i(\boldsymbol{y}) &\geq \eta(\boldsymbol{x},\boldsymbol{y})^T \nabla_{\boldsymbol{y}} g_i(\boldsymbol{y}),\ i\in I_{\mathcal{X}}(\bar{\boldsymbol{x}}),\\
h_i(\boldsymbol{x}) - h_i(\boldsymbol{y}) &\geq \eta(\boldsymbol{x},\boldsymbol{y})^T \nabla_{\boldsymbol{y}} h_i(\boldsymbol{y}),\ i\in I^+(\bar{\boldsymbol{x}})\cup I^+_{\mathcal{X}}(\bar{\boldsymbol{x}}),\\
h_i(\boldsymbol{y}) - h_i(\boldsymbol{x}) &\geq -\eta(\boldsymbol{x},\boldsymbol{y})^T \nabla_{\boldsymbol{y}} h_i(\boldsymbol{y}),\ i\in I^-(\bar{\boldsymbol{x}})\cup I^-_{\mathcal{X}}(\bar{\boldsymbol{x}}),
\end{aligned} \tag{1.47}$$

must have a solution $\eta(\boldsymbol{x},\boldsymbol{y})$. Denote Γ the matrix of the gradients, i.e.

$$\Gamma = \left(\nabla_{\boldsymbol{y}} f(\boldsymbol{y}), \nabla_{\boldsymbol{y}} g_{ij}(\boldsymbol{y}), \nabla_{\boldsymbol{y}} g_i(\boldsymbol{y}), \nabla_{\boldsymbol{y}} h_i(\boldsymbol{y}), -\nabla_{\boldsymbol{y}} h_i(\boldsymbol{y})\right)^T,$$

and $\boldsymbol{\lambda}$ the vector of the left-hand side. Then the matrix notation of the system (1.47) is as follows

$$\Gamma\eta(\boldsymbol{x},\boldsymbol{y}) \leq \boldsymbol{\lambda}. \tag{1.48}$$

From Gale's theorem of the alternative, vide [16], either the system $\Gamma\eta(\boldsymbol{x},\boldsymbol{y}) \leq \boldsymbol{\lambda}$ has a solution $\eta(\boldsymbol{x},\boldsymbol{y})$, or the system $\Gamma^T\boldsymbol{z} = 0$, $\boldsymbol{\lambda}^T\boldsymbol{z} = -1$, $\boldsymbol{z} \geq 0$, has a solution $\boldsymbol{z}$, but not both. Hence, one could in principle determine the existence of a kernel function by determining the nonexistence of a $\boldsymbol{z}$ in the latter system of equations, cf. [18].

1.3.3 Examples of exact penalty functions

The absolute value penalty function

The absolute value penalty function for (OP) is also a special case of L_q penalty function (1.5) for $q = 1$. Hence, it is sometimes referred to as the L_1 penalty function. It is given by

$$p_1(\boldsymbol{x}) = \|\max\{\mathbf{0},\boldsymbol{g}(\boldsymbol{x})\}\|_1 + \|\boldsymbol{h}(\boldsymbol{x})\|_1 = \sum_{i=1}^{m}\max\{0,g_i(\boldsymbol{x})\} + \sum_{i=1}^{l}|h_i(\boldsymbol{x})|. \tag{1.49}$$

The absolute value penalty function (1.49) is a commonly used exact penalty function. Since the dual norm to the L_1 norm is the L_∞ norm, cf. [19], it suffices to choose a penalty parameter

$$\mu > \max\left\{\bar{u}_1, \ldots, \bar{u}_m, |\bar{v}_1|, \ldots, |\bar{v}_l|\right\}$$

in order to satisfy the condition on the penalty parameter of theorems 1.2 or 1.3.

The L_∞ penalty function

The L_∞ penalty function is derived from the L_∞ norm (1.4), and therefore it is defined by

$$p_\infty(\boldsymbol{x}) = \max_{i=1,\ldots,m} \max\{0, g_i(\boldsymbol{x})\} + \max_{i=1,\ldots,l} |h_i(\boldsymbol{x})| \,. \tag{1.50}$$

Since the dual norm to the L_∞ norm is the L_1 norm, cf. [19], it suffices to choose a penalty parameter

$$\mu > \max\left\{\sum_{i=1}^{m} \bar{u}_i, \sum_{i=1}^{l} |\bar{v}_i|\right\}$$

in order to satisfy the condition on the penalty parameter of theorems 1.2 or 1.3.

Chapter 2

One-stage stochastic programming problems

By incorporating uncertainty into the original problem (OP), one-stage problem with uncertainty can be formulated as follows

$$\begin{array}{rl} \text{“ min ”} & f(\boldsymbol{x}, \boldsymbol{\omega}) \\ \text{“subject to”} & \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{\omega}) \leq \mathbf{0} \\ & \boldsymbol{x} \in \mathcal{X}, \end{array} \tag{PwUC}$$

where $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_s)^T$ is an s dimensional real-valued random vector defined on a probability space $(\Omega, \mathcal{F}, P)$ with known probability measure P and $\Omega \subset \mathbb{R}^s$, and $\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{\omega}) = (g_1(\boldsymbol{x}, \boldsymbol{\omega}), \ldots, g_m(\boldsymbol{x}, \boldsymbol{\omega}))^T$. All the functions occurring in the problem are defined on $\mathcal{X} \times \Omega$ and take values in the extended real numbers, i.e. $f(\boldsymbol{x}, \boldsymbol{\omega}) : \mathcal{X} \times \Omega \to \bar{\mathbb{R}}$, $g_i(\boldsymbol{x}, \boldsymbol{\omega}) : \mathcal{X} \times \Omega \to \bar{\mathbb{R}}$, $i = 1 \ldots, m$, where $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$ denotes the extended reals. Note that equality constraints, i.e. $\boldsymbol{h}(\boldsymbol{x}, \boldsymbol{\omega}) = \mathbf{0}$, are not consider in our formulation of the problem.

Since $\boldsymbol{\omega}$ is unknown, it remains a question how to interpret and solve the problem (PwUC). Considering the fact that the decision has to be made before $\boldsymbol{\omega}$ eventuate, a suitable reformulation of the problem (PwUC) is mandatory. As it was stated in the Introduction, the following two approaches has been eminently employed to solve the vagueness of the problem (PwUC) with respect to the treatment of the random vector $\boldsymbol{\omega}$:

- expected violation (or shortfall) penalty models,
- probabilistic (or chance constrained) programs.

We shall briefly outline the expected violation penalty technique and focus our attention on the probabilistic programs.

2.1 Expected violation penalty models

The expected violation penalty model formulation of the stochastic programs is as follows

$$\begin{array}{rl} \min & E\, F(\boldsymbol{x}, \boldsymbol{\omega}) \\ \text{subject to} & \boldsymbol{x} \in \mathcal{X}, \end{array} \tag{EVPM}$$

where $\mathcal{X}$ is a nonempty, closed set of decisions $\boldsymbol{x}$ which is completely independent of the probability measure, and $F(\boldsymbol{x}, \boldsymbol{\omega}) : \mathcal{X} \times \Omega \to \bar{\mathbb{R}}$ is a suitably chosen function which penalises the lost or cost caused by the decision $\boldsymbol{x} \in \mathcal{X}$ when the observation $\boldsymbol{\omega}$ occurs. The requirements on the function $F(\boldsymbol{x}, \boldsymbol{\omega})$ consists of measurability in $\boldsymbol{\omega}$ for every fixed $\boldsymbol{x} \in \mathcal{X}$ and well-defined expected value $E\,F(\boldsymbol{x}, \boldsymbol{\omega})$ for each $\boldsymbol{x} \in \mathcal{X}$. The objective function of the expected violation penalty problem (EVPM) may be rather complicated, e.g. multi-stage models in [33].

Let us show that the above formulation of the stochastic programming problem (PwUC) is truly based on the notion of optimising the total lost or cost on average. If the process repeats itself then by the Law of Large Numbers the average of the total cost will converge a.s. to the expectation $E\,F(\boldsymbol{x}, \boldsymbol{\omega})$, and indeed the solution of the expected violation penalty problem (EVPM) will be optimal on average, cf. [34].

One possible reformulation of the problem (PwUC) into an expected violation penalty program (EVPM) is the following. Suppose that $f(\boldsymbol{x}, \boldsymbol{\omega})$ and $\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{\omega})$ are measurable in $\boldsymbol{\omega}$ for every fixed $\boldsymbol{x} \in \mathcal{X}$ and the expected values $E\,f(\boldsymbol{x}, \boldsymbol{\omega})$, $E \max\{\mathbf{0}, \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{\omega})\}$ are well-defined for each $\boldsymbol{x} \in \mathcal{X}$. Then, the expected violation penalty program (EVPM) can be formulated as

$$\begin{aligned} \min \quad & E\left(f(\boldsymbol{x}, \boldsymbol{\omega}) + \boldsymbol{c}^T \max\{\mathbf{0}, \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{\omega})\}\right) \\ \text{subject to} \quad & \boldsymbol{x} \in \mathcal{X}, \end{aligned} \tag{2.1}$$

Note that the problem (2.1) is in the scalarised of the multi-objective problem

$$\begin{aligned} \min \quad & \{E\,f(\boldsymbol{x}, \boldsymbol{\omega}), E \max\{\mathbf{0}, \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{\omega})\}\} \\ \text{subject to} \quad & \boldsymbol{x} \in \mathcal{X}. \end{aligned} \tag{2.1-MO}$$

Therefore, if $\mathcal{X}$ is nonempty, convex and compact and the functions $E\,f(\boldsymbol{x}, \boldsymbol{\omega})$ and $E \max\{\mathbf{0}, \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{\omega})\}$ are convex in $\boldsymbol{x}$ then for every non-negative $\boldsymbol{c} \in \mathbb{R}^m$, $\boldsymbol{c} \neq \mathbf{0}$, every solution to the problem (2.1) is an efficient solution to the multi-objective problem (2.1-MO). If the functions $f(\boldsymbol{x}, \boldsymbol{\omega})$ and $\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{\omega})$ are convex in $\boldsymbol{x}$ then $E\,f(\boldsymbol{x}, \boldsymbol{\omega})$ and $E \max\{\mathbf{0}, \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{\omega})\}$ are convex in $\boldsymbol{x}$ as well. Let us show this for $E\,f(\boldsymbol{x}, \boldsymbol{\omega})$. For all $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{X}$ and $\alpha \in [0, 1]$, the following hold true

$$\begin{aligned} E\,f(\alpha \boldsymbol{x}_1 + (1-\alpha)\boldsymbol{x}_2, \boldsymbol{\omega}) &\leq E\,(f(\alpha \boldsymbol{x}_1, \boldsymbol{\omega}) + f((1-\alpha)\boldsymbol{x}_2, \boldsymbol{\omega})) \\ &\leq E\,f(\alpha \boldsymbol{x}_1, \boldsymbol{\omega}) + E\,f((1-\alpha)\boldsymbol{x}_2, \boldsymbol{\omega}). \end{aligned}$$

Thus, $E\,f(\boldsymbol{x}, \boldsymbol{\omega})$ is convex in $\boldsymbol{x}$. The convexity of $E \max\{\mathbf{0}, \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{\omega})\}$ in $\boldsymbol{x}$ is analogous.

Having noted that the problem (2.1) is a multi-objective program (2.1-MO), an other method, namely, the ϵ-constrained approach can be used to solve the problem. In this case, the problem (2.1-MO) is of the form

$$\begin{aligned} \min \quad & E\,f(\boldsymbol{x}, \boldsymbol{\omega}) \\ \text{subject to} \quad & E \max\{\mathbf{0}, \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{\omega})\} \leq \boldsymbol{\epsilon} \\ & \boldsymbol{x} \in \mathcal{X}, \end{aligned} \tag{2.1-ϵ-c}$$

where $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_m)^T$.

2.2 Probabilistic programming

Probabilistic (or chance constrained) programming problems were first introduced in [10]. In such problems, the set of feasible solutions $\mathcal{X}(P) \subset \mathbb{R}^n$ depends on the probability distribution P of the random part. The general formulation of probabilistic programs is as follows

$$\begin{array}{rl} \min & E\, f(\boldsymbol{x}, \boldsymbol{\omega}) \\ \text{subject to} & \boldsymbol{x} \in \mathcal{X}(P) . \end{array} \tag{CC}$$

Solving chance constrained problems is not an easy matter. In general, the structure of the set of feasible solutions $\mathcal{X}(P)$ is not convex, and hence, further specification are provided.

A common formulation of $\mathcal{X}(P)$ is derived from requirements on the reliability on the optimal solution. In general, one may require the random constraints $\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{\omega})$ to be meet for a prescribed probability thresholds $(1-\varepsilon)$ jointly or individually, depending on the nature of the problem. The reliability requirement $\varepsilon \in [0, 1]$ is chosen by the decision maker. Typically, it is rather small.

In the joint case, $\mathcal{X}(P) = \mathcal{X} \cap \mathcal{X}_\varepsilon(P)$, where $\mathcal{X}_\varepsilon(P)$ is defined as

$$\mathcal{X}_\varepsilon(P) = \{\boldsymbol{x} \in \mathbb{R}^n \colon\ P(\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{\omega}) \leq \boldsymbol{0}) \geq 1-\varepsilon\} \tag{2.2}$$

and called *joint probability constraints*. The structure of $\mathcal{X}_\varepsilon(P)$ may not be easy even in this particular case. To obtain a convex set, special assumptions on the probability distribution P are required. If the functions $g_i(\boldsymbol{x}, \boldsymbol{y})$, $i = 1, \ldots, m$, are quasi-convex jointly in both arguments and $\boldsymbol{\omega}$ is a random vector that has an α-concave probability distribution, cf. [12], then $\mathcal{X}_\varepsilon(P)$ is convex and closed. For more general results and details concerning joint probability constraints, see e.g. [12] or [28].

In the individual case, the structure of $\mathcal{X}_\varepsilon(P)$ is slightly easier due to separate treatment of $g_i(\boldsymbol{x}, \boldsymbol{\omega}) \leq 0$. Denote the probability thresholds ε_i on the ith constraint $g_i(\boldsymbol{x}, \boldsymbol{\omega}) \leq 0$, $i = 1, \ldots, m$. For this case, $\mathcal{X}_\varepsilon(P)$ is defined as

$$\mathcal{X}_\varepsilon(P) = \{\boldsymbol{x} \in \mathbb{R}^n \colon\ P(g_i(\boldsymbol{x}, \boldsymbol{\omega}) \leq \boldsymbol{0}) \geq 1-\varepsilon_i,\ i = 1, \ldots, m\} \tag{2.3}$$

and called *individual probability constraints*. If $s = m$ and ω_i, $i = 1, \ldots, m$, constitute the right hand sides of the constraints, i.e. $g_i(\boldsymbol{x}, \boldsymbol{\omega}) = g_i(x) - \omega_i$, the probabilistic constraints (2.3) become

$$P(g_i(\boldsymbol{x}) \leq \omega_i) \geq 1-\varepsilon_i,\ i = 1, \ldots, m$$

which is equivalent to

$$P(\omega_i \leq g_i(\boldsymbol{x})) \leq \varepsilon_i,\ i = 1, \ldots, m$$

and

$$g_i(\boldsymbol{x}) \leq u_{\varepsilon_i}(P_i),\ i = 1, \ldots, m,$$

where $u_{\varepsilon_i}(P_i)$ denotes the ε_i quantile of the marginal probability distribution of ω_i. Therefore, for quasi-convex $g_i(\boldsymbol{x})$, $i = 1, \ldots, m$, the set $X_\varepsilon(P)$ is convex, vide [3, Theorem 3.5.2]. For more general results and details concerning individual probability constraints, see e.g. [12].

Note that $P\left(\boldsymbol{g}\left(\boldsymbol{x},\boldsymbol{\omega}\right) \leq \mathbf{0}\right) = E\, I\left(\boldsymbol{g}\left(\boldsymbol{x},\boldsymbol{\omega}\right) \leq \mathbf{0}\right)$, and the chance constrained problem (CC) can be formulated as

$$\begin{aligned} \min \quad & E\, f\left(\boldsymbol{x},\boldsymbol{\omega}\right) \\ \text{subject to} \quad & E\, I\left(\boldsymbol{g}\left(\boldsymbol{x},\boldsymbol{\omega}\right) \leq \mathbf{0}\right) \geq 1-\varepsilon \\ & \boldsymbol{x} \in \mathcal{X}, \end{aligned} \tag{CC-ϵ-c}$$

which in the form of the ϵ-constrained approach of solving the multi-objective program

$$\begin{aligned} \min \quad & \left\{E\, f\left(\boldsymbol{x},\boldsymbol{\omega}\right), -E\, I\left(\boldsymbol{g}\left(\boldsymbol{x},\boldsymbol{\omega}\right) \leq \mathbf{0}\right)\right\} \\ \text{subject to} \quad & \boldsymbol{x} \in \mathcal{X}. \end{aligned} \tag{CC-MO}$$

Hence, for nonempty, convex, compact $\mathcal{X}$, convex $E\, f\left(\boldsymbol{x},\boldsymbol{\omega}\right)$ in $\boldsymbol{x}$ and concave $E\, I\left(\boldsymbol{g}\left(\boldsymbol{x},\boldsymbol{\omega}\right) \leq \mathbf{0}\right)$ in $\boldsymbol{x}$ there exists an $N \geq 0$ such that the optimal solutions to the probabilistic programming problem (CC) assuming joint probability constraints (2.2) and $\varepsilon \in [0,1]$, can be found by solving the scalarised form of the multi-objective problem (CC-MO)

$$\begin{aligned} \min \quad & E\left(f\left(\boldsymbol{x},\boldsymbol{\omega}\right) - N \cdot I\left(\boldsymbol{g}\left(\boldsymbol{x},\boldsymbol{\omega}\right) \leq \mathbf{0}\right)\right) \\ \text{subject to} \quad & \boldsymbol{x} \in \mathcal{X}. \end{aligned} \tag{CC-sc}$$

Note that the scalarised form (CC-sc) is an expected violation penalty problem (EVPM) with $F\left(\boldsymbol{x},\boldsymbol{\omega}\right) = f\left(\boldsymbol{x},\boldsymbol{\omega}\right) - N \cdot I\left(\boldsymbol{g}\left(\boldsymbol{x},\boldsymbol{\omega}\right) \leq \mathbf{0}\right)$, cf. [9].

These two aforementioned types of stochastic programming are not rivals but rather supplements and can be further combined in order to enhance the reliability of the optimal decision, cf. [9]. As we mentioned above, the solution of the expected violation penalty model optimises the lost or cost on average, whereas the probabilistic programming capture the reliability requirements or risk restrictions on the optimal solution. Hence, a "mixed" model is beneficial in such a way that one is able to control both the level of reliability and the extent of penalisation. The combination of the two approaches was first put forward by Prékopa [27].

> Prékopa [27] "... we are convinced that the best way of operating a stochastic system is to operate it with a prescribed (high) reliability and at the same time use penalties to punish discrepancies."

An example of a "mixed" model and its properties were recently presented in [5].

Chapter 3

Penalty function methods for stochastic programming

In section 2.2, we showed that a chance constrained problem (CC) can be reformulated into an expected violation penalty problem (EVPM), in which one employs a suitably chosen penalty function. Therefore, one can have a stochastic program (CC) with a fixed feasible region. Nevertheless, this reformulation is practically unsatisfactory. The deficiency of the scalarised form (CC-sc) of the multi-objective problem (CC-MO) associated with the probabilistic program (CC) is usually caused by relatively difficult evaluation of the objective function.

The idea of constructing a second program employing absolute value penalty function with a fixed set $\mathcal{X}$ of feasible solutions associated with the chance constrained problem (CC) with joint probability constraints (2.2) was proposed in [31]. They proposed to assign penalties $N \max\{0, g_i(\boldsymbol{x}, \boldsymbol{\omega})\}$ with positive penalty parameter N and to solve

$$\begin{aligned} \min \quad & E\, f(\boldsymbol{x}, \boldsymbol{\omega}) + N \sum_{i=1}^{m} E \max\{0, g_i(\boldsymbol{x}, \boldsymbol{\omega})\} \\ \text{subject to} \quad & \boldsymbol{x} \in \mathcal{X} \end{aligned} \tag{3.1}$$

instead of the chance constrained problem (CC). It is important to note that the two problems (CC) and (3.1) are not equivalent as the equivalent reformulation of the chance constrained problem (CC) using penalty function is the scalarised form (CC-sc), cf. section 2.2. However, one may expect that the two problems are asymptotically equivalent, i.e. for sufficiently large penalty parameter N there exists a reliability requirement $\varepsilon > 0$ such that the obtained solution to the penalty function problem (3.1) satisfies the joint probability constraints (2.2).

A rigorous proof of the asymptotic equivalence of chance constrained programs (CC) with one chance constraint and penalty function problems of form (3.1) is due to [15]. The approach was extended to an entire class of penalty function, cf. [9], and to problems with several individual and joint chance constraints, cf. [7]. The assumption of the theorems restricted their validity only to continuous probability distribution. The approach was recently extended to finite discrete probability distributions and exact penalty function methods as well, cf. [8].

3.1 Continuous probability distributions

First, we present a theorem of asymptotic equivalence which holds for continuous probability distributions. Prior to that, the formulations of the considered problems are provided.

3.1.1 The fomulation of multiple chance constrained and penalty function problems

Let $\mathcal{X}$ be a nonempty set in $\mathbb{R}^n$, $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_s)^T$ be a real-valued random vector defined on a probability space $(\Omega, \mathcal{F}, P)$ with known probability measure P and $\Omega \subset \mathbb{R}^s$, $f(x) : \mathbb{R}^n \to \mathbb{R}$ and $g_{ij}(\boldsymbol{x}, \boldsymbol{\omega}) : \mathbb{R}^n \times \mathbb{R}^s \to \mathbb{R}$, $j = 0, \ldots, m_i$, $i = 1, \ldots, k$, be real valued functions measurable in $\boldsymbol{\omega}$ for all $\boldsymbol{x} \in \mathcal{X}$, and denote $\boldsymbol{g}_i(\boldsymbol{x}, \boldsymbol{\omega}) = (g_{i1}(\boldsymbol{x}, \boldsymbol{\omega}), \ldots, g_{im_i}(\boldsymbol{x}, \boldsymbol{\omega}))^T$, $i = 1, \ldots, k$. Then the multiple chance constrained problem can be formulated as

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & P(\boldsymbol{g}_i(\boldsymbol{x}, \boldsymbol{\omega}) \leq \mathbf{0}) \geq 1 - \varepsilon_i, \; i = 1, \ldots, k, \\ & \boldsymbol{x} \in \mathcal{X}, \end{aligned} \tag{MCC}$$

where $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_k)^T$ is the vector of given reliability requirements $\varepsilon_i \in (0, 1)$ for the joint constraint $\boldsymbol{g}_i(\boldsymbol{x}, \boldsymbol{\omega})$, $i = 1, \ldots, k$. Denote $\psi_{\boldsymbol{\varepsilon}}$ the optimal value and $\boldsymbol{x}_{\boldsymbol{\varepsilon}}$ the optimal solution to the multiple chance constrained problem (MCC) for given reliability requirements $\boldsymbol{\varepsilon}$.

Note that the formulation of the multiple chance constrained problem (MCC) enables us to work with slightly generalised settings in comparison with probabilistic programming (CC) introduced in section 2.2. Since multiple reliability requirements on various constraints $\boldsymbol{g}_i(\boldsymbol{x}, \boldsymbol{\omega})$ can be set, it allows the decision maker to possess greater freedom of choice.

Consider the penalty functions $\phi_i(\boldsymbol{y}) : \mathbb{R}^{m_i} \to \mathbb{R}$, $i = 1, \ldots, k$, which are continuous and nondecreasing in their components, equal to 0 for $\boldsymbol{y} \leq \mathbf{0}$ and positive otherwise. Additionally, we denote

$$p_i(\boldsymbol{x}, \boldsymbol{\omega}) = \phi_i(g_{i1}(\boldsymbol{x}, \boldsymbol{\omega}), \ldots, g_{im_i}(\boldsymbol{x}, \boldsymbol{\omega})) : \mathbb{R}^n \times \mathbb{R}^s \to \mathbb{R}, \; i = 1, \ldots, k$$

the penalised constraints, cf. [7]. From the construction of $p_j(\boldsymbol{x}, \boldsymbol{\omega})$, the following relation holds true

$$P(\boldsymbol{g}_i(\boldsymbol{x}, \boldsymbol{\omega}) \leq \mathbf{0}) \geq 1 - \varepsilon_j \Longleftrightarrow P(p_i(\boldsymbol{x}, \boldsymbol{\omega}) > 0) \leq \varepsilon_i \tag{3.2}$$

for all $i = 1, \ldots, k$, cf. [7].

The penalty function problem associated with the multiple chance constrained problem (MCC) can be formulated as follows

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) + \mu \sum_{i=1}^{k} E\, p_i(\boldsymbol{x}, \boldsymbol{\omega}) \\ \text{subject to} \quad & \boldsymbol{x} \in \mathcal{X}, \end{aligned} \tag{P-MCC}$$

where μ is a positive penalty parameter. Denote φ_μ the optimal value and $\boldsymbol{x}_\mu$ the optimal solution to the penalty function problem (P-MCC) for a given penalty parameter μ.

3.1.2 The asymptotic equivalence of the two problems

Let us examine the the asymptotic equivalence of the two problems (MCC) and (P-MCC) under continuous probability distribution. The following theorem states the asymptotic equivalence in generalised settings, i.e. for arbitrary penalty function. The theorem is one of the main results in [7].

Theorem 3.1. *Consider the multiple chance constrained problems* (MCC) *and the corresponding penalty function problem* (P-MCC) *and assume that* $\mathcal{X} \subset \mathbb{R}^n$ *is nonempty, compact,* $f(\boldsymbol{x})$ *is continuous,*

(i) $g_{ij}(\boldsymbol{x}, \boldsymbol{\omega})$, $j = 1, \ldots, m_i$, $i = 1, \ldots, k$, *are almost surely continuous,*

(ii) there exists a nonnegative random variable $C(\boldsymbol{\omega})$ *with* $E\, C^{1+\kappa}(\boldsymbol{\omega}) < \infty$ *for some* $\kappa > 0$ *such that* $|p_i(\boldsymbol{x}, \boldsymbol{\omega})| \leq C(\boldsymbol{\omega})$, $i = 1, \ldots, k$, *for all* $\boldsymbol{x} \in \mathcal{X}$,

(iii) $E\, p_i(\boldsymbol{x}', \boldsymbol{\omega}) = 0$, $i = 1, \ldots, k$, *for some* $\boldsymbol{x}' \in \mathcal{X}$,

(iv) $P(g_{ij}(\boldsymbol{x}, \boldsymbol{\omega}) = 0) = 0$, $j = 1, \ldots, m_i$, $i = 1, \ldots, k$, *for all* $\boldsymbol{x} \in \mathcal{X}$.

Denote $\eta = \frac{\kappa}{2(1+\kappa)}$ *and for arbitrary* $\mu > 0$ *and* $\boldsymbol{\varepsilon} \in (0,1)^m$ *put*

$$\begin{aligned}
\varepsilon_j(\boldsymbol{x}) &= P(p_j(\boldsymbol{x}, \boldsymbol{\omega}) > 0), \ j = 1, \ldots, m, \\
\alpha_\mu(\boldsymbol{x}) &= \mu \sum_{j=1}^{m} E\, p_j(\boldsymbol{x}, \boldsymbol{\omega}), \\
\beta_{\boldsymbol{\varepsilon}}(\boldsymbol{x}) &= \varepsilon_{\max}^{-\eta} \sum_{j=1}^{m} E\, p_j(\boldsymbol{x}, \boldsymbol{\omega}),
\end{aligned}$$

where $\varepsilon_{\max}$ *denotes the maximum of the vector of the reliability requirements* $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_k)^T$, *and* $\left[1/\mu^{1/\eta}\right] = \left(1/\mu^{1/\eta}, \ldots, 1/\mu^{1/\eta}\right)^T$ *is a vector of length* k.

Then for any prescribed $\boldsymbol{\varepsilon} \in (0,1)^k$ *there always exists a sufficiently large* μ *so that the minimisation of the penalty function problem* (P-MCC) *generates optimal solutions* $\boldsymbol{x}_\mu$ *which also satisfy the multiple chance constraints* (MCC) *with the given* $\boldsymbol{\varepsilon}$.

Moreover, bounds on the optimal value $\psi_{\boldsymbol{\varepsilon}}$ *based on the optimal value* φ_μ *and vice versa can be constructed as*

$$\begin{aligned}
\varphi_{1/\varepsilon^{\eta}_{\max}(\boldsymbol{x}_\mu)} - \beta_{\boldsymbol{\varepsilon}(\boldsymbol{x}_\mu)}\left(\boldsymbol{x}_{\boldsymbol{\varepsilon}(\boldsymbol{x}_\mu)}\right) \leq \psi_{\boldsymbol{\varepsilon}(\boldsymbol{x}_\mu)} &\leq \varphi_\mu - \alpha_\mu(\boldsymbol{x}_\mu) \\
\psi_{\boldsymbol{\varepsilon}(\boldsymbol{x}_\mu)} + \alpha_\mu(\boldsymbol{x}_\mu) \leq \varphi_\mu &\leq \psi_{\left[1/\mu^{1/\eta}\right]} + \beta_{\left[1/\mu^{1/\eta}\right]}\left(\boldsymbol{x}_{\left[1/\mu^{1/\eta}\right]}\right),
\end{aligned} \tag{3.3}$$

with

$$\lim_{\mu\to\infty} \alpha_\mu(\boldsymbol{x}_\mu) = \lim_{\mu\to\infty} \varepsilon_j(\boldsymbol{x}_\mu) = \lim_{\varepsilon_{\max}\to 0+} \beta_{\boldsymbol{\varepsilon}}(\boldsymbol{x}_{\boldsymbol{\varepsilon}}) = 0$$

for any sequences of optimal solutions $\boldsymbol{x}_\mu$ *and* $\boldsymbol{x}_{\boldsymbol{\varepsilon}}$.

Proof. The proof is based on Theorem 1.1 and can be found in [7]. □

Note that Theorem 3.1 does not make any statement on the convergence of optimal solutions. It merely relates the optimal values for certain reliability requirements $\boldsymbol{\varepsilon}$ and the penalty parameter μ, cf. [7].

Since the only points of discontinuity of the probability function

$$P\left(g_{ij}\left(\boldsymbol{x},\boldsymbol{\omega}\right)\leq 0,\ j=1,\ldots,m_i\right),$$

in the decision variable $\boldsymbol{x}$ is $g_{ij}\left(\boldsymbol{x},\boldsymbol{\omega}\right)=0$, $j=1,\ldots,m_i$, $i=1,\ldots,k$, assumption (iv) ensures its continuity for any $\boldsymbol{x}\in\mathcal{X}$, cf. [7].

The bounds (3.3) and the terms $\varepsilon_i\left(\boldsymbol{x}\right)$, $i=1,\ldots,k$, $\alpha_\mu\left(\boldsymbol{x}\right)$ and $\beta_{\boldsymbol{\varepsilon}}\left(\boldsymbol{x}\right)$ depend on the chosen penalty functions $\phi_i\left(\boldsymbol{y}\right)$, $i=1,\ldots,k$. Nonetheless, numeric evaluation of the bounds may cause difficulties. One cannot compute $\beta_{\boldsymbol{\varepsilon}(\boldsymbol{x}_\mu)}\left(\boldsymbol{x}_{\boldsymbol{\varepsilon}(\boldsymbol{x}_\mu)}\right)$ without having the optimal solution $\boldsymbol{x}_{\boldsymbol{\varepsilon}(\boldsymbol{x}_\mu)}$, which is not aimed to be found or may be extremely difficult to obtain, cf. [7].

Let us examine assumption (iii). Since $p_i\left(\boldsymbol{x},\boldsymbol{\omega}\right)\geq 0$ a.s. for all $\boldsymbol{x}\in\mathcal{X}$ and $E\,p_i\left(\boldsymbol{x}',\boldsymbol{\omega}\right)=0$ for some $\boldsymbol{x}'\in\mathcal{X}$, it follows that $p_i\left(\boldsymbol{x}',\boldsymbol{\omega}\right)=0$ a.s. This assumption states that $\boldsymbol{x}'$ is a permanently feasible solution, i.e. a solution to the multiple chance constrained problem (MCC) with reliability requirements $\boldsymbol{\varepsilon}=\boldsymbol{0}$. Therefore, assumption (iii) can be severely restrictive. In general, the overall feasible set may shrink with increasing reliability requirements $\boldsymbol{\varepsilon}$ to the empty set, and hence, Theorem 3.1 may fail for probability measures with an unbounded support, cf. [9].

In section 1.3, we discussed that if $p\left(\boldsymbol{x}_\mu\right)=0$ for some $\mu>0$, then $\boldsymbol{x}_\mu$ is an optimal solution to the original problem (OP). In this case it means that if $E\,p_i\left(\boldsymbol{x}_\mu,\boldsymbol{\omega}\right)=0$, $i=1,\ldots,k$ for some $\mu>0$ then $\boldsymbol{x}_\mu$ is a permanently feasible solution.

3.2 Finite discrete probability distributions

Under discrete probability distribution with finite number of realisations, the theorem of asymptotic equivalence requires less assumptions. In addition, one can employ exact penalty function methods, section 1.3, in order to obtain the equivalence of the problems.

3.2.1 The fomulation of chance constrained and penalty function problems

Let the probability distribution of the random vector $\boldsymbol{\omega}$ be discrete with finite number of realisations (scenarios) $\boldsymbol{\omega}^s$, $s=1,\ldots,S$, and denote $0<\pi_s<1$, $s=1,\ldots,S$, $\sum_{s=1}^S\pi_s=1$ the associated probabilities. For such probability distributions, the probabilistic program (MCC) can be formulated as

$$\begin{aligned}\min\quad & f\left(\boldsymbol{x}\right)\\ \text{subject to}\quad & \sum_{s=1}^{S}\pi_s\, I\left(\boldsymbol{g}_i\left(\boldsymbol{x},\boldsymbol{\omega}^s\right)\leq\boldsymbol{0}\right)\geq 1-\varepsilon_i,\ i=1,\ldots,k\\ & \boldsymbol{x}\in\mathcal{X},\end{aligned}\tag{MCCd}$$

and the corresponding penalty function problem (P-MCC) can be formulated as

$$\begin{aligned}\min\quad & f\left(\boldsymbol{x}\right)+\mu\sum_{i=1}^{k}\sum_{s=1}^{S}\pi_s p_i\left(\boldsymbol{x},\boldsymbol{\omega}^s\right)\\ \text{subject to}\quad & \boldsymbol{x}\in\mathcal{X}.\end{aligned}\tag{P-MCCd}$$

The notations of the optimal value and the optimal solution to the multiple chance constrained problem (MCCd) and the corresponding penalty function problem (P-MCCd) do not change.

3.2.2 The asymptotic equivalence of the two problems

First, we examine the the asymptotic equivalence of the two problems (MCCd) and (P-MCCd) under finite discrete probability distribution. The following theorem is analogous to Theorem 3.1, and thus it states the asymptotic equivalence in generalised settings, i.e. for arbitrary penalty function. It also extends Theorem 1 in [8] to multiple chance constrained problems.

Theorem 3.2. *Consider the multiple chance constrained problem* (MCCd) *and the corresponding penalty function problem* (P-MCCd) *and assume that $\mathcal{X} \subset \mathbb{R}^n$ is a nonempty, compact set, $f(\boldsymbol{x})$ a continuous function,*

(i) $g_{ij}(\boldsymbol{x}, \boldsymbol{\omega}^s)$, $j = 1, \ldots, m_i$, $i = 1, \ldots, k$, *are continuous for all* $s = 1, \ldots, S$,

(ii) $g_{ij}(\boldsymbol{x}, \boldsymbol{\omega}^s) \leq 0$ *for all* $j = 1, \ldots, m_i$, $i = 1, \ldots, k$, $s = 1, \ldots, S$ *and at least one* $\boldsymbol{x}' \in \mathcal{X}$.

For arbitrary $\gamma \in (0,1)$, $\mu > 0$ and $\boldsymbol{\varepsilon} \in (0,1)^k$ put

$$\begin{aligned}
\varepsilon_i(\boldsymbol{x}) &= \sum_{s=1}^{S} \pi_s \, I\left(p_i(\boldsymbol{x}, \boldsymbol{\omega}^s) > 0\right), \; i = 1, \ldots, k, \\
\alpha_\mu(\boldsymbol{x}) &= \mu \sum_{i=1}^{k} \sum_{s=1}^{S} \pi_s p_i(\boldsymbol{x}, \boldsymbol{\omega}^s), \\
\beta_{\boldsymbol{\varepsilon}}(\boldsymbol{x}) &= \varepsilon_{\max}^{-\gamma} \sum_{i=1}^{k} \sum_{s=1}^{S} \pi_s p_i(\boldsymbol{x}, \boldsymbol{\omega}^s),
\end{aligned}$$

where $\varepsilon_{\max}$ denotes the maximum of the vector of the reliability requirements $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_k)^T$ and $\left[1/\mu^{1/\eta}\right] = \left(1/\mu^{1/\eta}, \ldots, 1/\mu^{1/\eta}\right)^T$ is a vector of length k.

Then for any prescribed $\boldsymbol{\varepsilon} \in (0,1)^k$ there always exists a sufficiently large μ so that the minimisation of the penalty function problem (P-MCCd) *generates optimal solutions $\boldsymbol{x}_\mu$ which also satisfy the multiple chance constraints* (MCCd) *with the given reliability requirements $\boldsymbol{\varepsilon}$.*

Moreover, bounds on the optimal value $\psi_{\boldsymbol{\varepsilon}}$ based on the optimal value φ_μ and vice versa can be constructed as

$$\begin{aligned}
\varphi_{1/\varepsilon_{\max}^{\gamma}(\boldsymbol{x}_\mu)} - \beta_{\boldsymbol{\varepsilon}(\boldsymbol{x}_\mu)}\left(\boldsymbol{x}_{\boldsymbol{\varepsilon}(\boldsymbol{x}_\mu)}\right) \leq \;& \psi_{\boldsymbol{\varepsilon}(\boldsymbol{x}_\mu)} &&\leq \; \varphi_\mu - \alpha_\mu(\boldsymbol{x}_\mu), \\
\psi_{\boldsymbol{\varepsilon}(\boldsymbol{x}_\mu)} + \alpha_\mu(\boldsymbol{x}_\mu) \leq \;& \varphi_\mu &&\leq \; \psi_{[1/\mu^{1/\gamma}]} + \beta_{[1/\mu^{1/\gamma}]}\left(\boldsymbol{x}_{[1/\mu^{1/\gamma}]}\right),
\end{aligned} \tag{3.4}$$

with

$$\lim_{\mu \to \infty} \alpha_\mu(\boldsymbol{x}_\mu) = \lim_{\varepsilon_{\max} \to 0+} \beta_{\boldsymbol{\varepsilon}}(\boldsymbol{x}_{\boldsymbol{\varepsilon}}) = \lim_{\mu \to \infty} \varepsilon_i(\boldsymbol{x}_\mu) = 0, \; i = 1, \ldots, k$$

for any sequences of optimal solutions $\boldsymbol{x}_\mu$ and $\boldsymbol{x}_{\boldsymbol{\varepsilon}}$.

Proof. Assumption (ii) implies that for every vector $\boldsymbol{\varepsilon} > 0$ there exists a $\boldsymbol{x}_{\boldsymbol{\varepsilon}} \in \mathcal{X}$ such that

$$\sum_{s=1}^{S} \pi_s \, I\left(g_{i1}\left(\boldsymbol{x}_{\boldsymbol{\varepsilon}}, \boldsymbol{\omega}^s\right) \leq 0, \ldots, g_{im_i}\left(\boldsymbol{x}_{\boldsymbol{\varepsilon}}, \boldsymbol{\omega}^s\right) \leq 0\right) \geq 1 - \varepsilon_i, \ i = 1, \ldots, k.$$

Denote

$$C = \max_{i=1,\ldots,k} \max_{s=1,\ldots,S} \max_{\boldsymbol{x}\in\mathcal{X}} p_i\left(\boldsymbol{x}, \boldsymbol{\omega}^s\right),$$

which is finite due to our assumptions. Then for any $\boldsymbol{\varepsilon} > 0$ the following relations hold

$$\sum_{i=1}^{k}\sum_{s=1}^{S} \pi_s p_i\left(\boldsymbol{x}, \boldsymbol{\omega}^s\right) \leq C \sum_{i=1}^{k}\sum_{s=1}^{S} \pi_s \, I\left(p_i\left(\boldsymbol{x}, \boldsymbol{\omega}^s\right) > 0\right) \leq kC\varepsilon_{\max}.$$

Hence, for $\varepsilon_{\max} \to 0+$ and arbitrary $\gamma \in (0,1)$

$$\beta_{\boldsymbol{\varepsilon}}\left(\boldsymbol{x}\right) = \varepsilon_{\max}^{-\gamma} \sum_{i=1}^{k}\sum_{s=1}^{S} \pi_s p_i\left(\boldsymbol{x}, \boldsymbol{\omega}^s\right) \leq kC\varepsilon_{\max}^{1-\gamma} \to 0. \tag{3.5}$$

We denote

$$\delta_\mu = \sum_{i=1}^{k}\sum_{s=1}^{S} \pi_s p_i\left(\boldsymbol{x}_\mu, \boldsymbol{\omega}^s\right)$$

for a sequence $\boldsymbol{x}_\mu$ of optimal solutions to the penalty function problem (P-MCCd). Our assumptions and the general properties of the penalty function method, vide Theorem 1.1, ensures that $\delta_\mu \to 0+$ and moreover $\alpha_\mu\left(\boldsymbol{x}_\mu\right) = \mu\delta_\mu \to 0$ as $\mu \to \infty$. Let $G_i\left(\boldsymbol{x}, \cdot\right)$ denote the distribution function of $p_i\left(\boldsymbol{x}, \boldsymbol{\omega}\right)$ for a fixed $\boldsymbol{x}$ defined by

$$G_i\left(\boldsymbol{x}, y\right) = P\left(p_i\left(\boldsymbol{x}, \boldsymbol{\omega}\right) \leq y\right) = \sum_{s=1}^{S} \pi_s \, I\left(p_i\left(\boldsymbol{x}, \boldsymbol{\omega}\right) \leq y\right), \ i = 1, \ldots, k.$$

The we can rewrite $\varepsilon_i\left(\boldsymbol{x}_\mu\right)$, $i = 1, \ldots, k$, as

$$\begin{aligned}
\varepsilon_i\left(\boldsymbol{x}_\mu\right) &= \sum_{s=1}^{S} \pi_s \, I\left(p_i\left(\boldsymbol{x}_\mu, \boldsymbol{\omega}^s\right) > 0\right) \\
&= \sum_{s=1}^{S} \pi_s \, I\left(0 < p_i\left(\boldsymbol{x}_\mu, \boldsymbol{\omega}^s\right) \leq \sqrt{\delta_\mu}\right) + \sum_{s=1}^{S} \pi_s \, I\left(p_i\left(\boldsymbol{x}_\mu, \boldsymbol{\omega}^s\right) > \sqrt{\delta_\mu}\right)
\end{aligned}$$

The first summand can be rewritten by using the empirical distribution function as

$$\sum_{s=1}^{S} \pi_s \, I\left(0 < p_i\left(\boldsymbol{x}_\mu, \boldsymbol{\omega}^s\right) \leq \sqrt{\delta_\mu}\right) = G_i\left(\boldsymbol{x}_\mu, \sqrt{\delta_\mu}\right) - G_i\left(\boldsymbol{x}_\mu, 0\right).$$

The second summand can be rewritten by using Markov's inequality

$$P\left(X > a\right) \leq \frac{E\,X}{a}$$

as

$$\sum_{s=1}^{S} \pi_s \, I\left(p_i\left(\boldsymbol{x}_\mu, \boldsymbol{\omega}^s\right) > \sqrt{\delta_\mu}\right) \leq \frac{1}{\sqrt{\delta_\mu}} \sum_{s=1}^{S} \pi_s p_i\left(\boldsymbol{x}_\mu, \boldsymbol{\omega}^s\right) = \sqrt{\delta_\mu}.$$

Therefore, for $i = 1, \dots, k$ we have

$$\varepsilon_i (\boldsymbol{x}_\mu) \le G_i \left(\boldsymbol{x}_\mu, \sqrt{\delta_\mu}\right) - G_i (\boldsymbol{x}_\mu, 0) + \sqrt{\delta_\mu} \to 0 \text{ as } \mu \to \infty \tag{3.6}$$

as the distribution function $G_i (\boldsymbol{x}, \cdot)$ is right continuous. This means that for μ large enough one can generate feasible solution for the chance constrained problem with arbitrary low reliability requirements $\boldsymbol{\varepsilon}$.

The optimal solution $\boldsymbol{x}_\mu$ of the penalty function problem (P-MCCd) with penalty parameter μ is obviously feasible for the multiple chance constrained problem (MCCd) with reliability requirements $\boldsymbol{\varepsilon} (\boldsymbol{x}_\mu) = (\varepsilon_1 (\boldsymbol{x}_\mu), \dots, \varepsilon_k (\boldsymbol{x}_\mu))^T$ as relation (3.2) holds with $\varepsilon_i = \varepsilon_i (\boldsymbol{x}_\mu)$. Therefore, we get the lower bound for the optimal value φ_μ of the penalty function problem (P-MCCd)

$$\begin{aligned}
\varphi_\mu &= f (\boldsymbol{x}_\mu) + \mu \sum_{i=1}^{k} \sum_{s=1}^{S} \pi_s p_i (\boldsymbol{x}_\mu, \boldsymbol{\omega}^s) \\
&\ge f \left(\boldsymbol{x}_{\boldsymbol{\varepsilon}(\boldsymbol{x}_\mu)}\right) + \mu \sum_{i=1}^{k} \sum_{s=1}^{S} \pi_s p_i (\boldsymbol{x}_\mu, \boldsymbol{\omega}^s) \\
&= \psi_{\boldsymbol{\varepsilon}(\boldsymbol{x}_\mu)} + \alpha_\mu (\boldsymbol{x}_\mu) .
\end{aligned}$$

This bound is directly employed in bounds (3.4).

Since for the optimal solution $\boldsymbol{x}_{\boldsymbol{\varepsilon}}$ to the multiple chance constrained problem (MCCd) holds

$$f (\boldsymbol{x}_{\boldsymbol{\varepsilon}}) + \varepsilon_{\max}^{-\gamma} \sum_{i=1}^{k} \sum_{s=1}^{S} \pi_s p_i (\boldsymbol{x}_{\boldsymbol{\varepsilon}}, \boldsymbol{\omega}^s) \ge \varphi_{\varepsilon_{\max}^{-\gamma}},$$

hence we the lower bound for the optimal value $\psi_{\boldsymbol{\varepsilon}}$ of the multiple chance constrained problem (MCCd)

$$\begin{aligned}
\psi_{\boldsymbol{\varepsilon}} &= \left(\psi_{\boldsymbol{\varepsilon}} + \varepsilon_{\max}^{-\gamma} \sum_{i=1}^{k} \sum_{s=1}^{S} \pi_s p_i (\boldsymbol{x}_{\boldsymbol{\varepsilon}}, \boldsymbol{\omega}^s)\right) - \varepsilon_{\max}^{-\gamma} \sum_{i=1}^{k} \sum_{s=1}^{S} \pi_s p_i (\boldsymbol{x}_{\boldsymbol{\varepsilon}}, \boldsymbol{\omega}^s) \\
&\ge \varphi_{\varepsilon_{\max}^{-\gamma}} - \varepsilon_{\max}^{-\gamma} \sum_{i=1}^{k} \sum_{s=1}^{S} \pi_s p_i (\boldsymbol{x}_{\boldsymbol{\varepsilon}}) = \varphi_{\varepsilon_{\max}^{-\gamma}} - \beta_{\boldsymbol{\varepsilon}} (\boldsymbol{x}_{\boldsymbol{\varepsilon}})
\end{aligned}$$

This bound is then employed in (3.4) by setting $\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon} (\boldsymbol{x}_\mu)$ and $\boldsymbol{\varepsilon} = \left[1/\mu^{1/\gamma}\right]$. This completes the proof. □

Assumption (ii) requires a permanently feasible solution to multiple constrained problem (MCCd), which is defined as the optimal solution to

$$\begin{aligned}
\min \quad & f (\boldsymbol{x}) \\
\text{subject to} \quad & \boldsymbol{g}_i (\boldsymbol{x}, \boldsymbol{\omega}^s) \le \boldsymbol{0}, \ i = 1, \dots, k, \ s = 1, \dots, S \\
& \boldsymbol{x} \in \mathcal{X}.
\end{aligned} \tag{PF-MCCd}$$

Note that for $\varepsilon_i < \min_{s=1,\dots,S} \pi_s, \ i = 1, \dots, k,$ any solution to the multiple chance constrained problem (MCCd) is a solution to the permanently feasible problem (PF-MCCd), and therefore is permanently feasible, cf. [8]. An analogous criterion to obtain a permanently feasible solution can be stated as $\sum_{i=1}^k \sum_{s=1}^S \pi_s p_i (\boldsymbol{x}_\mu, \boldsymbol{\omega}^s) = 0$ for some $\mu > 0$ and $\boldsymbol{x}_\mu \in \mathcal{X}$, cf. section 1.2.

3.2.3 The asymptotic equivalence of the two problems using exact penalisation

One of the disadvantages of exterior penalty function methods is the necessity of infinite increase of the penalty parameter. Thus, one must solve several problems for ascendant penalty parameters and estimate the reliability of the obtained solution to stochastic programming problems. Analogously to deterministic problems, the penalty function approach for stochastic programming can be further improved by means of exact penalty function methods.

In [8], an exact penalty function method using the absolute value penalty function under general assumptions was proposed.

The following theorem compounds the exact penalty function method for convex and concave functions stated by Theorem 1.3 and penalty function methods for stochastic programming under discrete probability distributions stated by Theorem 3.2. Since in Theorem 1.3 arbitrary vector norms can be theoretically used, the following theorem extends the usable penalty functions in case of convex functions.

Theorem 3.3. *Consider the two problems* (MCCd) *and* (P-MCCd) *and assume that* $\mathcal{X} = \{\boldsymbol{x} \in \mathbb{R}^n \colon g_i(\boldsymbol{x}) \leq 0,\ i = 1, \ldots, M,\ h_i(\boldsymbol{x}) = 0,\ i = 1, \ldots, L\}$ *is nonempty,* $\mathcal{C} \supset \mathcal{X}$ *is an open convex set,* $f(\boldsymbol{x})$ *is a convex function on* $\mathcal{C}$*,*

- *(i)* $g_{ij}(\boldsymbol{x}, \boldsymbol{\omega}^s)$, $j = 1, \ldots, m_i$, $i = 1, \ldots, k$ *and* $s = 1, \ldots, S$, *are convex functions on* $\mathcal{C}$*,*
- *(ii)* $g_i(\boldsymbol{x})$, $i = 1, \ldots, M$, *are convex functions on* $\mathcal{C}$*,*
- *(iii)* $h_i(\boldsymbol{x})$ *are affine for all* $i = 1, \ldots, L$*,*
- *(iv) there exists a KKT point for the permanently feasible problem* (PF-MCCd).

Let

$$p_i(\boldsymbol{x}, \boldsymbol{\omega}) = \left\| \max\left\{\mathbf{0}, \boldsymbol{g}_i(\boldsymbol{x}, \boldsymbol{\omega})\right\} \right\|,\ i = 1, \ldots, k,$$

where $\|\cdot\|$ *is any fixed vector norm in* $\mathbb{R}^{m_i}$, $i = 1, \ldots, k$, *respectively. For arbitrary* $\mu > 0$ *and* $\boldsymbol{\varepsilon} \in (0,1)^k$ *put*

$$\begin{aligned}
\varepsilon_i(\boldsymbol{x}) &= \sum_{s=1}^{S} \pi_s I\left(p_i(\boldsymbol{x}, \boldsymbol{\omega}^s) > 0\right),\ i = 1, \ldots, k \\
\alpha_\mu(\boldsymbol{x}) &= \mu \sum_{i=1}^{k} \sum_{s=1}^{S} \pi_s p_i(\boldsymbol{x}, \boldsymbol{\omega}^s), \\
\beta_{\boldsymbol{\varepsilon}}(\boldsymbol{x}) &= \varepsilon_{\max}^{-1} \sum_{i=1}^{k} \sum_{s=1}^{S} \pi_s p_i(\boldsymbol{x}, \boldsymbol{\omega}^s),
\end{aligned}$$

where $\varepsilon_{\max}$ *denotes the maximum of the vector of reliabilities* $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_k)^T$.

Then there exists a $\tilde{\mu}$ *that any optimal solution to the penalty function problem* (P-MCCd) *is also an optimal solution to the permanently feasible problem* (PF-MCCd).

In addition, suppose that for every $\mu \leq \tilde{\mu}$ *there exists an optimal solution* $\boldsymbol{x}_\mu$ *to the penalty function problem* (P-MCCd) *and for every* $\boldsymbol{\varepsilon} \in [\tilde{\varepsilon}, 1)^k$, $\tilde{\varepsilon} < \min_{s=1,\ldots,S} \pi_s$, *there exists an optimal solution* $\boldsymbol{x}_{\boldsymbol{\varepsilon}}$ *to the multiple chance constrained problem* (MCCd).

Then for any prescribed $\boldsymbol{\varepsilon} \in (0,1)^k$ *there always exists* $\mu \leq \tilde{\mu}$ *so that the optimal solution* $\boldsymbol{x}_\mu$ *satisfies the chance constraints with the given* $\boldsymbol{\varepsilon}$.

Moreover, bounds on the optimal value $\psi_{\boldsymbol{\varepsilon}}$ *based on the optimal value* φ_μ *and vice versa can be constructed as*

$$\begin{aligned} \varphi_{1/\varepsilon_{\max}(\boldsymbol{x}_\mu)} - \beta_{\boldsymbol{\varepsilon}(\boldsymbol{x}_\mu)}\left(\boldsymbol{x}_{\boldsymbol{\varepsilon}(\boldsymbol{x}_\mu)}\right) \leq \ & \psi_{\boldsymbol{\varepsilon}(\boldsymbol{x}_\mu)} & \leq \ & \varphi_\mu - \alpha_\mu\left(\boldsymbol{x}_\mu\right), \\ \psi_{\boldsymbol{\varepsilon}(\boldsymbol{x}_\mu)} + \alpha_\mu\left(\boldsymbol{x}_\mu\right) \leq \ & \varphi_\mu & \leq \ & \psi_{[1/\mu]} + \beta_{[1/\mu]}\left(\boldsymbol{x}_{[1/\mu]}\right), \end{aligned} \tag{3.7}$$

with

$$\lim_{\mu \to \tilde{\mu}-} \alpha_\mu\left(\boldsymbol{x}_\mu\right) = \lim_{\varepsilon_{\max} \to \tilde{\varepsilon}+} \beta_{\boldsymbol{\varepsilon}}\left(\boldsymbol{x}_{\boldsymbol{\varepsilon}}\right) = \lim_{\mu \to \tilde{\mu}-} \varepsilon_i\left(\boldsymbol{x}_\mu\right) = 0, \ i = 1, \ldots, k$$

for any sequences of optimal solutions $\boldsymbol{x}_\mu$ *and* $\boldsymbol{x}_{\boldsymbol{\varepsilon}}$, *where* $[1/\mu] = (1/\mu, \ldots, 1/\mu)^T$ *is a vector of length* k.

Proof. The proof is analogous to the proof of Theorem 3.2, and hence, we shall discuss only the validity of the main steps.

Note the permanently feasible problem (PF-MCCd) can be considered as a problem with multiple inequality constraints (MCOP). Its corresponding exact penalty function problem can be formulated as

$$\begin{aligned} \min \quad & f\left(\boldsymbol{x}\right) + \sum_{i=1}^{k} \sum_{s=1}^{S} \mu_i^s p_i\left(\boldsymbol{x}, \boldsymbol{\omega}^s\right) \\ \text{subject to} \quad & \boldsymbol{x} \in \mathcal{X}. \end{aligned} \tag{3.8}$$

Obviously, the formulation of the penalty function problem (P-MCCd) differs from problem (3.8) in the penalty parameters. In the penalty function problem (P-MCCd), a common penalty parameter is multiplied by the probabilities of the scenarios, i.e. $\mu_i^s = \mu \pi_s$.

It follows from the assumption and Theorem 1.3 that there exist penalty parameters $\tilde{\mu}_i^s$, $i = 1 \ldots, k$, $s = 1, \ldots, S$, so that any solution to problem (3.8) is an optimal solution to the permanently feasible problem (PF-MCCd). Then for a common penalty parameter such that

$$\tilde{\mu} \pi_s \geq \tilde{\mu}_i^s, \ i = 1, \ldots, k, \ s = 1, \ldots, S,$$

which can be written as

$$\tilde{\mu} \geq \max_{s=1,\ldots,S} \left\{ \frac{\max\left\{\tilde{\mu}_1^s, \ldots, \tilde{\mu}_k^s\right\}}{\pi_s} \right\},$$

any solution to the penalty function problem (P-MCCd) is an optimal solution to the permanently feasible problem (PF-MCCd), and therefore

$$\delta_{\tilde{\mu}} = \sum_{i=1}^{k} \sum_{s=1}^{S} \pi_s p_i\left(\boldsymbol{x}_{\tilde{\mu}}, \boldsymbol{\omega}^s\right) = 0.$$

Since $p_i(\cdot, \boldsymbol{\omega})$, $i = 1, \ldots, k$, is continuous for fixed $\boldsymbol{\omega}$, we have

$$\lim_{\mu \to \tilde{\mu}-} \delta_\mu = \sum_{i=1}^{k} \sum_{s=1}^{S} \pi_s \lim_{\mu \to \tilde{\mu}-} p_i(\boldsymbol{x}_\mu, \boldsymbol{\omega}^s) = \sum_{i=1}^{k} \sum_{s=1}^{S} \pi_s p_i(\boldsymbol{x}_{\tilde{\mu}}, \boldsymbol{\omega}^s) = 0$$

and

$$\lim_{\mu \to \tilde{\mu}-} \alpha_\mu(\boldsymbol{x}_\mu) = \lim_{\mu \to \tilde{\mu}-} \mu \delta_\mu = \tilde{\mu} \delta_{\tilde{\mu}} = 0.$$

Analogously to (3.6), we have

$$\varepsilon(\boldsymbol{x}_\mu) \le G\left(\boldsymbol{x}_\mu, \sqrt{\delta_\mu}\right) - G(\boldsymbol{x}_\mu, 0) + \sqrt{\delta_\mu} \to 0 \text{ as } \mu \to \tilde{\mu}-$$

Therefore, for μ large enough one can generate a feasible solution for the chance constrained problem with arbitrary low ε.

Since any solution $\boldsymbol{x}_{\boldsymbol{\varepsilon}}$, where $\varepsilon_i \le \tilde{\varepsilon}$, $i = 1, \ldots, k$, is a solution to the permanently feasible problem (PF-MCCd), and thus $p_i(\boldsymbol{x}_{\boldsymbol{\varepsilon}}, \omega^s) = 0$ for all $i = 1, \ldots, k$ and $s = 1, \ldots, S$. Therefore, we can write

$$\lim_{\varepsilon_{\max} \to \tilde{\varepsilon}+} \beta_{\boldsymbol{\varepsilon}}(\boldsymbol{x}_\varepsilon) = \varepsilon_{\max}^{-1} \sum_{i=1}^{k} \sum_{s=1}^{S} \pi_s \lim_{\varepsilon_{\max} \to \tilde{\varepsilon}+} p_i(\boldsymbol{x}_{\boldsymbol{\varepsilon}}, \boldsymbol{\omega}^s) = 0.$$

The bounds (3.7) do not change and therefore remain valid. This completes the proof. □

Analogously to section 1.3, one can employ invex functions with respect to the same $\eta(\boldsymbol{x}, \boldsymbol{y})$ in lieu of convex functions in Theorem 3.3. Note that in such cases, differentiable functions are required to exist.

Theorem 3.4. *Consider the two problems* (MCCd) *and* (P-MCCd) *and assume that* $\mathcal{X} = \{\boldsymbol{x} \in \mathbb{R}^n : g_i(\boldsymbol{x}) \le 0,\ i = 1, \ldots, M,\ h_i(\boldsymbol{x}) = 0,\ i = 1, \ldots, L\}$ *is nonempty,* $\mathcal{O} \supset \mathcal{X}$ *is an open set,* $f(\boldsymbol{x})$ *is invex on* $\mathcal{O}$ *with respect to an* $\eta(\boldsymbol{x}, \boldsymbol{y})$,

(i) $g_{ij}(\boldsymbol{x}, \boldsymbol{\omega}^s)$, $j = 1, \ldots, m_i$, $i = 1, \ldots, k$ *and* $s = 1, \ldots, S$, *are invex on* $\mathcal{O}$ *with respect to the same* $\eta(\boldsymbol{x}, \boldsymbol{y})$,

(ii) $g_i(\boldsymbol{x})$, $i = 1, \ldots, M$, *are invex on* $\mathcal{O}$ *with respect to the same* $\eta(\boldsymbol{x}, \boldsymbol{y})$,

(iii) $h_i(\boldsymbol{x})$ *are affine for all* $i = 1, \ldots, L$,

(iv) there exists a KKT point for the permanently feasible problem (PF-MCCd).

Let

$$p_i(\boldsymbol{x}, \boldsymbol{\omega}) = \|\max\{\mathbf{0}, \boldsymbol{g}_i(\boldsymbol{x}, \boldsymbol{\omega})\}\|,\ i = 1, \ldots, k,$$

where $\|\cdot\|$ *is any fixed vector norm in* $\mathbb{R}^{m_i}$, $i = 1, \ldots, k$, *respectively. For arbitrary* $\mu > 0$ *and* $\boldsymbol{\varepsilon} \in (0,1)^k$ *put*

$$\varepsilon_i(\boldsymbol{x}) = \sum_{s=1}^{S} \pi_s I(p_i(\boldsymbol{x}, \boldsymbol{\omega}^s) > 0),\ i = 1, \ldots, k$$

$$\alpha_\mu(\boldsymbol{x}) = \mu \sum_{i=1}^{k} \sum_{s=1}^{S} \pi_s p_i(\boldsymbol{x}, \boldsymbol{\omega}^s),$$

$$\beta_{\boldsymbol{\varepsilon}}(\boldsymbol{x}) = \varepsilon_{\max}^{-1} \sum_{i=1}^{k} \sum_{s=1}^{S} \pi_s p_i(\boldsymbol{x}, \boldsymbol{\omega}^s),$$

where $\varepsilon_{\max}$ denotes the maximum of the vector of reliabilities $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_k)^T$.

Then there exists a $\tilde{\mu}$ that any optimal solution to the penalty function problem (P-MCCd) *is also an optimal solution to the permanently feasible problem* (PF-MCCd)*.*

In addition, suppose that for every $\mu \leq \tilde{\mu}$ there exists an optimal solution $\boldsymbol{x}_\mu$ to the penalty function problem (P-MCCd) *and for every $\boldsymbol{\varepsilon} \in [\tilde{\varepsilon}, 1)^k$, $\tilde{\varepsilon} < \min_{s=1,\ldots,S} \pi_s$, there exists an optimal solution $\boldsymbol{x}_{\boldsymbol{\varepsilon}}$ to the multiple chance constrained problem* (MCCd)*.*

Then for any prescribed $\boldsymbol{\varepsilon} \in (0,1)^k$ there always exists $\mu \leq \tilde{\mu}$ so that the optimal solution $\boldsymbol{x}_\mu$ satisfies the chance constraints with the given $\boldsymbol{\varepsilon}$.

Moreover, bounds on the optimal value $\psi_{\boldsymbol{\varepsilon}}$ based on the optimal value φ_μ and vice versa can be constructed as

$$\begin{aligned} \varphi_{1/\varepsilon_{\max}(\boldsymbol{x}_\mu)} - \beta_{\boldsymbol{\varepsilon}(\boldsymbol{x}_\mu)}\left(\boldsymbol{x}_{\boldsymbol{\varepsilon}(\boldsymbol{x}_\mu)}\right) \leq \ & \psi_{\boldsymbol{\varepsilon}(\boldsymbol{x}_\mu)} && \leq \varphi_\mu - \alpha_\mu\left(\boldsymbol{x}_\mu\right), \\ \psi_{\boldsymbol{\varepsilon}(\boldsymbol{x}_\mu)} + \alpha_\mu\left(\boldsymbol{x}_\mu\right) \leq \ & \varphi_\mu && \leq \psi_{[1/\mu]} + \beta_{[1/\mu]}\left(\boldsymbol{x}_{[1/\mu]}\right), \end{aligned} \tag{3.9}$$

with

$$\lim_{\mu \to \tilde{\mu}-} \alpha_\mu\left(\boldsymbol{x}_\mu\right) = \lim_{\varepsilon_{\max} \to \tilde{\varepsilon}+} \beta_{\boldsymbol{\varepsilon}}\left(\boldsymbol{x}_{\boldsymbol{\varepsilon}}\right) = \lim_{\mu \to \tilde{\mu}-} \varepsilon_i\left(\boldsymbol{x}_\mu\right) = 0, \ i = 1, \ldots, k$$

for any sequences of optimal solutions $\boldsymbol{x}_\mu$ and $\boldsymbol{x}_{\boldsymbol{\varepsilon}}$, where $[1/\mu] = (1/\mu, \ldots, 1/\mu)^T$ is a vector of length k.

Proof. The proof is analogous to the proof of Theorem 3.3. It differs only in applying Theorem 1.4 in lieu of Theorem 1.3. This completes the proof. $\square$

Remark that having $\mathcal{X}$ compact is sufficient for the existence of $\boldsymbol{x}_\mu$ for $\mu \leq \tilde{\mu}$ and $\boldsymbol{x}_{\boldsymbol{\varepsilon}}$ for $\boldsymbol{\varepsilon} \in [\tilde{\varepsilon}, 1)^k$. In many practical cases, having $\mathcal{X}$ compact is not severely restrictive since the variables are usually bounded.

If one wants to obtain exact convergence of the bounds (3.7) on the optimal values, additional condition on μ is required to be set. Since $1/\mu$ is the upper bound on φ_μ has to converge to $\tilde{\varepsilon}$, it follows that $\mu \geq \max\{\tilde{\mu}, 1/\tilde{\varepsilon}\}$, cf. [8].

In both cases, assumption (iv) can be restrictive for certain practical problems. Not only is a permanently feasible solution required, but a KKT point must exist as well. In order for a minimum point to satisfy the conditions of being a KKT point, the problem must satisfy some constraint qualifications, which can be found for example in [3, chapter 5].

Theoretically, one is able to evaluate the threshold $\tilde{\mu}$ defined in Theorem 1.3 and Theorem 1.4. Denote $\bar{\boldsymbol{x}}$ a KKT point for the permanently feasible problem (PF-MCCd) associated with $g_{ij}(\bar{\boldsymbol{x}}, \boldsymbol{\omega}^s)$, $j = 1, \ldots, m_i$, $i = 1, \ldots, k$ and $s = 1, \ldots, S$ and $\bar{\boldsymbol{u}}_i^s = \left(\bar{u}_{i1}^s, \ldots, \bar{u}_{im_i}^s\right)^T$ the vector of these Lagrangian multipliers, $i = 1, \ldots, k$, $s = 1, \ldots, S$. Applying the notation of either Theorem 1.3 or Theorem 1.4, we can write

$$\bar{\mu}_i^s = \|\bar{\boldsymbol{u}}_i^s\|', \ i = 1, \ldots, k, \ s = 1, \ldots, S.$$

By using the notation of the proof of Theorem 3.3 and the requirements on the penalty parameters $\tilde{\mu}_i^s$ stated by either Theorem 1.3 or Theorem 1.4, we obtain

$$\tilde{\mu}_i^s > \bar{\mu}_i^s, \ i = 1, \ldots, k, \ s = 1, \ldots, S.$$

Since

$$\tilde{\mu} \geq \max_{s=1,\ldots,S} \left\{ \frac{\max\left\{\tilde{\mu}_1^s, \ldots, \tilde{\mu}_k^s\right\}}{\pi_s} \right\},$$

the following holds inequality holds for the threshold

$$\tilde{\mu} > \max_{s=1,\ldots,S} \left\{ \frac{\max\left\{\|\bar{\boldsymbol{u}}_1^s\|', \ldots, \|\bar{\boldsymbol{u}}_k^s\|'\right\}}{\pi_s} \right\}. \tag{3.10}$$

Note that the constraints in $\mathcal{X}$ are not penalised, and therefore they do not have influence on the threshold, cf. Theorem 1.3 and Theorem 1.4.

Chapter 4

Numerical study

In order to demonstrate the practical usability of the stated theorems, in particular Theorem 3.3 and Theorem 3.4, we shall conduct a numerical study on solving a VaR-constrained portfolio selection problem [2]. Prior to further specification of the illustrative problem, we shall briefly present the definition of the Value-at-Risk (VaR), delineate the VaR-constrained portfolio selection problem and introduce a mixed-integer reformulation approach to solve chance constrained problems.

4.1 Value-at-Risk

Define the random loss $L(\boldsymbol{x}, \boldsymbol{\omega})$ dependent on a decision $\boldsymbol{x} \in \mathcal{X} \subset \mathbb{R}^n$ (weights of assets in our portfolio, portfolio allocation) and a random rates of returns $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_m)^T$, which is a real-valued random vector defined on a probability space $(\Omega, \mathcal{F}, P)$.

Definiton 4.1 (Value-at-Risk (VaR), [30])**.** The Value-at-Risk of a portfolio allocated by $\boldsymbol{x}$ at $(1-\varepsilon)\cdot 100\%$ confidence level is defined by

$$\mathrm{VaR}_{1-\varepsilon}(\boldsymbol{x}) = \min\left\{\xi\colon\ P\left(L(\boldsymbol{x}, \boldsymbol{\omega}) \le \xi\right) \ge 1-\varepsilon\right\}.$$

Note that $R(\boldsymbol{x}, \boldsymbol{\omega}) = -L(\boldsymbol{x}, \boldsymbol{\omega})$ is the random return on the portfolio allocated by a decision $\boldsymbol{x}$.

4.2 VaR-constrained portfolio selection problem

As the designation of the problem may foretell, the VaR-constrained portfolio selection problem does not attempt to optimise VaR. Actually, it is indirectly constrained by the following probability constraint

$$P\left(\boldsymbol{x}^T\boldsymbol{\omega} \ge R_d\right) \ge 1-\varepsilon, \tag{4.1}$$

where $\boldsymbol{\omega}$ is the random vector of the rates of returns on n risky assets, $\boldsymbol{x}$ is the vector of weights of assets in our portfolio (portfolio allocation), R_d is a desired minimal rate of return on the portfolio, and $\varepsilon \in [0,1]$ is the reliability requirement.

Let us demonstrate that the constraint (4.1) is truly a VaR constraint. Denote L the random loss of the portfolio. Then, it evidently holds that $L = -\boldsymbol{x}^T\boldsymbol{\omega}$. By substituting L in (4.1) we get

$$P\left(L \leq -R_d\right) \geq 1 - \varepsilon.$$

Having considered Definition 4.1, we get $\mathrm{VaR}_{1-\varepsilon}(\boldsymbol{x}) \leq -R_d$, i.e. we showed that the VaR of the portfolio allocated by $\boldsymbol{x}$ at a $(1-\varepsilon)\cdot 100\%$ confidence level is constrained by $-R_d$.

In general, VaR-constrained portfolio selection problem is formulated as

$$\begin{aligned} \min \quad & \boldsymbol{x}^T \Sigma \boldsymbol{x} \\ \text{s.t.} \quad & P\left(\boldsymbol{x}^T\boldsymbol{\omega} \geq R_d\right) \geq 1 - \varepsilon \\ & \boldsymbol{x}^T\boldsymbol{e} = 1 \\ & \boldsymbol{x} \geq \boldsymbol{0} \end{aligned} \tag{VaRcPSP}$$

where Σ is the sample covariance of the rates of returns, and $\boldsymbol{e} = (1, \ldots, 1)^T$ is a vector of length n.

The VaR-constrained portfolio selection problem (VaRcPSP) can be perceived as a special type of mean-variance problems (or the Markowitz model [13]), in which the constraint

$$E\,\boldsymbol{x}^T\boldsymbol{\omega} \geq R_d$$

is replaced by the chance constraint (4.1). Hence, the objective function of the VaR-constrained portfolio selection problem (VaRcPSP) is quadratic, and the function in the chance constraint, formally $g\left(\boldsymbol{x}, \boldsymbol{\omega}\right) = R_d - \boldsymbol{x}^t\boldsymbol{\omega}$, is affine.

4.3 Mixed-integer reformulation of chance constraint problems

Consider the multiple chance constrained problem under finite discrete probability distribution (MCCd). Suppose that $\mathcal{X}$ is compact and $g_{ij}\left(\boldsymbol{x}, \boldsymbol{\omega}^s\right)$ is continuous for every triples (i, j, s). In this case, the multiple chance constrained problem (MCCd) can be reformulated as a large mixed-integer nonlinear program, cf. [29],

$$\begin{aligned} \min \quad & f\left(\boldsymbol{x}\right) \\ \text{subject to} \quad & \boldsymbol{g}_i\left(\boldsymbol{x}, \boldsymbol{\omega}^s\right) \leq \; M\left(1 - u_{is}\right)\boldsymbol{e}, \; i = 1, \ldots, k, \; s = 1, \ldots, S \\ & \sum_{s=1}^{S} \pi_s u_{is} \geq 1 - \varepsilon_i, \; i = 1, \ldots, k \\ & \boldsymbol{x} \in \mathcal{X} \\ & u_{is} \in \{0, 1\}\,, \; i = 1, \ldots, k, \; s = 1, \ldots, S, \end{aligned} \tag{MI-MCCd}$$

where $M = \max_i \max_j \max_s \sup_{\boldsymbol{x}\in\mathcal{X}} g_{ij}\left(\boldsymbol{x}, \boldsymbol{\omega}^s\right)$.

Due to potential multitude of binary variables u_{is}, it may be difficult to solve the mixed-integer problem (MI-MCCd) even with solvers specifically designed for mixed-integer problems, cf. [8].

4.4 The solved problem

Suppose that $\boldsymbol{\omega}$ is a random vector with finite equiprobable realisations (scenarios) $\boldsymbol{\omega}^s$, $s = 1, \dots, S$, hence a associated probability π_s is equal to $\frac{1}{S}$. Additionally, suppose that any allocation can not exceed 0.5. In these settings, the VaR-constrained portfolio selection problem (VaRcPSP) is of the form

$$\begin{aligned} \min \quad & \boldsymbol{x}^T \Sigma \boldsymbol{x} \\ \text{s.t.} \quad & \frac{1}{S} \sum_{s=1}^{S} I\left(\boldsymbol{x}^T \boldsymbol{\omega}^s \geq R_d\right) \geq 1 - \varepsilon \\ & \boldsymbol{x}^T \boldsymbol{e} = 1 \\ & \mathbf{0.5} \geq \boldsymbol{x} \geq \mathbf{0}, \end{aligned} \tag{VaRcPSPd}$$

where $\mathbf{0.5} = (0.5, \dots, 0.5)^T$.

Let us present the mixed-integer and penalty function forms of the proposed problem (VaRcPSPd).

4.4.1 The mixed-integer form

The probability distribution independent feasible region $\mathcal{X}$ in this case is equal to $\{\boldsymbol{x} \in \mathbb{R}^n \colon \boldsymbol{x}^t \boldsymbol{e} = 1, \boldsymbol{x} \geq \mathbf{0}\}$ and is evidently compact. Since $g(\boldsymbol{x}, \boldsymbol{\omega}^s) = R_d - \boldsymbol{x}^T \boldsymbol{\omega}^s$, $s = 1, \dots, S$, are indubitably affine, they are continuous as well. Hence, we can apply the mixed-integer reformulation introduced in section 4.3. The mixed-integer reformulation of the VaR-constrained portfolio selection problem (VaRcPSP) is of the form

$$\begin{aligned} \min \quad & \boldsymbol{x}^T \Sigma \boldsymbol{x} \\ \text{s.t.} \quad & R_d - \boldsymbol{x}^t \boldsymbol{\omega}^s \leq M(1 - u_s), \; s = 1, \dots, S \\ & \frac{1}{S} \sum_{s=1}^{S} u_s \geq 1 - \varepsilon \\ & \boldsymbol{x}^T \boldsymbol{e} = 1 \\ & \mathbf{0.5} \geq \boldsymbol{x} \geq \mathbf{0} \\ & u_s \in \{0, 1\}, \; s = 1, \dots, S, \end{aligned} \tag{MI-VaRcPSPd}$$

where $M = \max_s \sup_{\boldsymbol{x} \in \mathcal{X}} (R_d - \boldsymbol{x}^t \boldsymbol{\omega}^s)$. Since there is merely one chance constraint employed, the total number of binary variables solely depends on the total number of scenarios.

4.4.2 The penalty function form

In every practical problem, it is utterly important to choose a suitable penalty function. A suitable penalty function must reflect the nature of the chance constraints, e.g. the L_∞ penalty function is typically used for the joint chance constraints (2.2), cf. [7]. Beside the L_∞ penalty function, the absolute value penalty function is commonly used as well, cf. [7].

Since there is only one function in the chance constraint of the stochastic programming problem (VaRcPSPd), the absolute value penalty function and the L_∞ penalty function coincide. By the reason of the prevalence of the mentioned penalty functions and their coincidence, we shall use the following penalty function

$$p(\boldsymbol{x}, \boldsymbol{\omega}) = \max\{0, g(\boldsymbol{x}, \boldsymbol{\omega})\}$$

in order to solve the VaR-constrained portfolio selection problem (VaRcPSPd). The corresponding penalty function problem is of the form

$$\begin{aligned} \min \quad & \boldsymbol{x}^T \Sigma \boldsymbol{x} + \frac{\mu}{S} \sum_{s=1}^{S} \max\left\{0, R_d - \boldsymbol{x}^T \boldsymbol{\omega}^s\right\} \\ \text{s.t.} \quad & \boldsymbol{x}^T e = 1 \\ & \mathbf{0.5} \geq \boldsymbol{x} \geq \mathbf{0}, \end{aligned} \tag{P-VaRcPSPd}$$

where $\mu > 0$ is the penalty parameter.

Let us examine the eligibility of the considered penalty function problem (P-VaRcPSPd) for applicability of Theorem 3.3 or Theorem 3.4. The objective function is quadratic, the function in the chance constraint is affine, and the remaining function are likewise affine. Hence, each and every function occurring in the problem is differentiable convex on $\mathbb{R}^n$, and the penalty function problem (P-VaRcPSPd) satisfies assumptions (i), (ii) and (iii) of Theorem 3.3. Since the sole chance constraint employed is affine, Kuhn-Tucker's constraint qualification holds true, and hence, the KKT conditions are necessary for optimality. As the objective function is continuous and the set of permanently feasible solutions

$$\mathcal{X}_{PF} = \left\{\boldsymbol{x} \in \mathbb{R}^n \colon \boldsymbol{x}^t \boldsymbol{e} = 1,\ \mathbf{0.5} \geq \boldsymbol{x} \geq \mathbf{0},\ R_d - \boldsymbol{x}^t \boldsymbol{\omega}^s \leq,\ s = 1, \ldots, S\right\} \tag{4.2}$$

is compact (any allocation can not exceed 1), $\mathcal{X}_{PF}$ must be nonempty in order for an optimal solution to exist. In such a case, the penalty function problem (VaRcPSPd) also satisfies assumption (iv).

Note that the penalty function problem (VaRcPSPd) itself is a convex program. The convexity of the constraints follows from the discussion above, as does the convexity of the first term of the objective function. Since every vector norm preserves convexity, and so does $\max\{0, \cdot\}$, the penalised constraint is convex as well.

By adding auxiliary variables, the penalty function problem (P-VaRcPSPd) can be reformulated into a quadratic programming problem. Let $\boldsymbol{y} = (y_1, \ldots, y_S)^T$ denote the auxiliary variables. Then the quadratic programming reformulation is as follows

$$\begin{aligned} \max \quad & -\boldsymbol{x}^T \Sigma \boldsymbol{x} - \frac{\mu}{S} \sum_{s=1}^{S} y_s \\ \text{s.t.} \quad & \boldsymbol{x}^T e = 1 \\ & \mathbf{0.5} \geq \boldsymbol{x} \geq \mathbf{0} \\ & \boldsymbol{x}^T \boldsymbol{\omega}^s + y_s \geq R_d,\ s = 1, \ldots, S \\ & \boldsymbol{y} \geq \mathbf{0}. \end{aligned} \tag{P-VaRcPSPdQP}$$

4.4.3 Data and parameters

For solving the two problems (MI-VaRcPSPd) and (P-VaRcPSPd), historical weekly rates of returns on every component of the Dow Jones Industrial Average were employed. The Dow Jones Industrial Average is a stock market index consisting of 30 major American companies representing various industries. The used time interval ranges from 05/01/2004 to 01/07/2013 amounting to 496 scenarios. As stated above, all the scenarios are supposed to be equiprobable.

In order for the permanently feasible region (4.2) to be nonempty, the desired minimal rate of return R_d must be carefully chosen. Allowing the minimal rate of return to be negative may increase the chance to obtain a nonempty permanently feasible region. Hence, the desired minimal rate of return was chosen to be

$$R_d = -0.1.$$

The confidence level for VaR typically equals to 95% or 99%. Hence, the reliability requirements of the chance constraint (4.1) were chosen to equal to these typical levels. Besides, three more reliability requirements were chosen for illustrative purposes. The reliability requirements are represented in Table 4.1.

ε_1	ε_2	ε_3	ε_4	ε_5
0.001	0.002	0.005	0.01	0.05

Table 4.1: Reliability requirements

In order to demonstrate the capability of the penalty function method to generate comparable solutions, the penalty parameters represented in Table 4.2 were chosen.

μ_1	μ_2	μ_3	μ_4	μ_5	μ_6	μ_7	μ_8	μ_9	μ_{10}
0.001	0.01	0.1	0.5	1	1.5	2	2.5	3	3.5

Table 4.2: Penalty parameters

4.4.4 Results

For solving the penalty function form (P-VaRcPSPd) numerically, its quadratic programming reformulation (P-VaRcPSPdQP) was used. For verifying the nonemptiness of the set of permanently feasible solutions, the modelling system GAMS was applied. This modelling system was likewise employed for solving the mixed-integer form (MI-VaRcPSPd) of the chance constrained problem and the quadratic programming form (P-VaRcPSPdQP) of the penalty function problem. In both cases, the Ilog CPLEX 12.5 solver was used.

Let us examine the numerical results. In Tables 4.3 and 4.4, the optimal values of both the chance constrained problems and the penalty function problems are summarised. Additionally, the reliabilities of the optimal solutions to the penalty function problems is illustrated in Table 4.4. The column "Reliability" was evaluated by substituting the optimal decisions into the VaR-constraint (4.1). Obviously, the penalty function problem (P-VaRcPSPd) is able to generate comparable solutions. From Table 4.4, it is also apparent that the threshold $\tilde{\mu}$ lies between 2.5 and 3.

ε	ψ_ε
0.001	0.000465
0.002	0.000306
0.005	0.000306
0.01	0.000306
0.05	0.000306

Table 4.3: Chance constrained problems - optimal values

μ	"Reliability"	φ_μ	μ	"Reliability"	φ_μ
0.001	0.99798	0.000306	1.5	0.99798	0.000426
0.01	0.99798	0.000307	2	0.99798	0.000448
0.1	0.99798	0.000317	2.5	0.99798	0.000461
0.5	0.99798	0.000355	3	1	0.000465
1	0.99798	0.000395	3.5	1	0.000465

Table 4.4: Penalty function problems - optimal values and "reliabilites"

In order to confirm that the threshold $\tilde{\mu}$ lies between 2.5 and 3, the Lagrangian multipliers associated with $g_i(\boldsymbol{x}, \boldsymbol{\omega}^s)$, $s = 1, \ldots, 496$ were obtained by solving the permanently feasible problem using GAMS and the Ilog CPLEX 12.5 solver. The largest obtained Lagrangian multiplier equalled approximately to 0.006. By substituting into the right-hand side of inequality (3.10), we obtained that the threshold $\tilde{\mu} > 2.976$. Hence, it accords with our previous conjecture.

The data and the GAMS programs solving the mixed-integer problem form (MI-VaRcPSPd), the quadratic programming form (P-VaRcPSPdQP) and the permanently feasible problem are available on the enclosed CD.

Conclusion

In this thesis, we dealt with penalty function methods for stochastic programming. It was shown that multiple chance constrained and the corresponding penalty function problems are asymptotically equivalent, i.e. by solving the corresponding penalty function problem, one can obtain highly reliable solutions to the original multiple chance constrained program, and vice versa. Thus, this equivalence provides an alternate approach to solve probabilistic programming problems, and its research was worthy of interest.

The aim of the thesis was to propose new penalty function methods for stochastic programming and to validate their usefulness on a numerical example. In order for this plan to be fulfilled, the following steps were done.

In chapter 1, exterior and exact penalty functions methods for deterministic nonlinear problems were studied. The main aim of this chapter was to put forward new exact penalty function methods. In order to fulfil this aim, a theorem concerning exact penalty function methods for convex programming problems using arbitrary vector norm as penalty function was proposed. Although this theorem let us use arbitrary vector norm as penalty function, extending it to other types of functions was likewise important. Thus, invex and incave functions were employed. These functions intend to generalise convex and concave functions by substituting the linear term $(\boldsymbol{x} - \boldsymbol{y})$ appearing in the definition of differentiable convex functions by an arbitrary vector-valued function $\eta(\boldsymbol{x}, \boldsymbol{y})$, called *kernel function*. This vector-valued function $\eta(\boldsymbol{x}, \boldsymbol{y})$ plays an inevitable role in the definition of invex and incave functions and may restrict the use of other types of functions in certain cases. The existence of such a common vector-valued function was discussed in subsection 1.3.2.

In chapter 2, a reformulation of probabilistic programming problems (CC) into expected violation penalty models (EVPM) was shown for certain functions occurring the problems. This reformulation is closely related to the notion of employing penalty functions methods for stochastic programming, and hence, it served as a first encounter with the topic.

In chapter 3, the main theorems of the thesis were put forward. Our attention was mostly focused on multiple chance constrained problems under finite discrete probability distributions. In subsection 3.2.2, we extended Theorem 1 in [8] to multiple chance constrained problems. In subsection 3.2.3, we proposed two analogous theorems using exact penalty function methods for convex and invex functions, respectively. Similar exact penalty function methods for stochastic programming were studied in [8]. Despite the fact that the proposed theorem in [8] uses a modified calmness property [8, Definition 1] which provides more general settings in comparison with convex and invex functions, respectively, the choice of the penalty function is not arbitrary. The theorem allows us to use only a modified absolute value penalty function. As the proposed theorems in subsection 3.2.3 are based on theorems in section 1.3, they enable us to use any vector norm as penalty function. Therefore, we pay for the variability with reduced generality.

In chapter 4, the capability of the penalty function methods to generate comparable solutions was demonstrated on a small illustrative example. It was also showed that the exact penalisation truly functioned since the optimal solutions to the penalty function problems with penalty parameter exceeding the threshold $\tilde{\mu}$ equaled to the optimal solution to the permanently feasible problem.

Even though we managed to extend exact penalty function methods to non-convex problems through invex functions with respect to the same kernel function, the importance of the requirement of having identical kernel function for each function occurring in the problems may diminish or even reverse their generalising effect in certain cases. Hence, future research of the topic should concern possible extension of the proposed methods to other types of functions, e.g. pseudo- and quasi-convex functions.

Other crucial requirement of the proposed theorems concerning exact penalty function methods for stochastic programming is the existence of a KKT point to the permanently feasible problem (PF-MCCd). The analogous theorem in [8] merely requires having a permanently feasible solution, which is naturally weaker condition than the existence of a KKT point. Thus, research of possible elimination of the named requirement could improve the applicability of the proposed theorems.

Unfortunately, the above mentioned enhancements were beyond the scope of this thesis. Nonetheless, these improvements may provide substantial extensions of penalty function methods for stochastic programming and therefore are worth of future research.

Appendix A

A.1 Vector norms

Definiton A.1 (dual norm)**.** Let $\|\cdot\|$ be a vector norm on $\mathbb{R}^n$. Then the dual norm $\|\cdot\|'$ of $\|\cdot\|$ is defined as follows

$$\|\boldsymbol{x}\|' = \sup_{\|\boldsymbol{y}\|=1} \boldsymbol{x}^T \boldsymbol{y}, \quad \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n.$$

Theorem A.1 (generalised Cauchy–Schwarz inequality)**.** *For any vector norm $\|\cdot\|$ on $\mathbb{R}^n$ and its corresponding dual norm $\|\cdot\|'$ the following inequality holds true*

$$\boldsymbol{x}^T \boldsymbol{y} \leq \|\boldsymbol{x}\| \cdot \|\boldsymbol{y}\|' \quad \textit{for all } \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n. \tag{A.1}$$

Proof. For $\boldsymbol{x} = \mathbf{0}$ or $\boldsymbol{y} = \mathbf{0}$ the inequality evidently holds. Suppose that $\boldsymbol{x} \neq \mathbf{0}$ and $\boldsymbol{y} \neq \mathbf{0}$. Then it follows from the definition above that

$$\|\boldsymbol{y}\|' = \sup_{\|\boldsymbol{z}\|=1} \boldsymbol{y}^T \boldsymbol{z} \geq \frac{\boldsymbol{y}^T \boldsymbol{x}}{\|\boldsymbol{x}\|}.$$

This completes the proof. □

A.2 Nonlinear programming

Definiton A.2 (affine functions)**.** A function $f(\boldsymbol{x}) : \mathbb{R}^n \to \mathbb{R}$ is called affine if it is of the form

$$f(\boldsymbol{x}) = \alpha + \boldsymbol{c}^t \boldsymbol{x},$$

where α is a scalar and $\boldsymbol{c}$ is a vector of length n.

Definiton A.3 (regular solution)**.** Let $\bar{\boldsymbol{x}}$ be a feasible solution to the original problem (OP), and denote $I(\bar{\boldsymbol{x}}) = \{i \in (1, \ldots, M) : g_i(\bar{\boldsymbol{x}}) = 0\}$. Suppose that $f(\boldsymbol{x})$ and $g_i(\boldsymbol{x})$, $i \in I(\boldsymbol{x})$, $h_i(\boldsymbol{x})$, $i = 1, \ldots, l$, are differentiable at $\bar{\boldsymbol{x}}$. Then $\bar{\boldsymbol{x}}$ is called regular solution to the problem if $\nabla_{\boldsymbol{x}} g_i(\bar{\boldsymbol{x}})$, $i \in I(\bar{x})$, and $\nabla_{\boldsymbol{x}} h_i(\bar{\boldsymbol{x}})$, $i = 1, \ldots, l$, are linearly independent.

Definiton A.4 (KKT point)**.** Let $\bar{\boldsymbol{x}}$ be feasible for the original problem (OP) and $g_i(\boldsymbol{x})$, $i = 1, \ldots, m$, $h_i(\boldsymbol{x})$, $i = 1 \ldots, l$ be differentiable at $\bar{\boldsymbol{x}}$. Then $\bar{\boldsymbol{x}}$ is called KKT point if there exist Lagrangian multipliers $\bar{u}_i$, $i = 1, \ldots, m$, and $\bar{v}_i$,

$i = 1, \dots, l$ associated with the inequality $g_i(\boldsymbol{x})$, $i = 1, \dots, m$, and the equality constraints $h_i(\boldsymbol{x})$, $i = 1, \dots, l$, respectively, such that

$$\begin{aligned} \nabla_{\boldsymbol{x}} f(\bar{\boldsymbol{x}}) + \sum_{i=1}^{m} \bar{u}_i \nabla_{\boldsymbol{x}} g_i(\bar{\boldsymbol{x}}) + \sum_{i=1}^{l} \bar{v}_i \nabla_{\boldsymbol{x}} h_i(\bar{\boldsymbol{x}}) &= \mathbf{0} \\ \bar{u}_i g_i(\bar{\boldsymbol{x}}) &= 0, \qquad i = 1, \dots, m \\ \bar{u}_i &\geq 0, \qquad i = 1, \dots, m. \end{aligned}$$

Note that the first equality is equivalent to

$$\nabla_{\boldsymbol{x}} \mathcal{L}(\bar{\boldsymbol{x}}; \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}) = \mathbf{0}.$$

Definiton A.5 (second-order sufficient conditions)**.** Let $\bar{\boldsymbol{x}}$ be a KKT point for the original problem (OP) and $g_i(\boldsymbol{x})$, $i = 1, \dots, m$, $h_i(\boldsymbol{x})$, $i = 1 \dots, l$ be twice differentiable at $\bar{\boldsymbol{x}}$. Then the second-order sufficient conditions hold true at $\bar{\boldsymbol{x}}$ if for all $\boldsymbol{d}$ in the critical cone

$$\begin{aligned} \mathcal{C} = \{\boldsymbol{d} \neq \mathbf{0} \colon \nabla_{\boldsymbol{x}} g_i(\bar{\boldsymbol{x}})^T \boldsymbol{d} &= 0 \quad \text{for} \quad i \in I(\bar{\boldsymbol{x}}), \bar{u}_i > 0, \\ \nabla_{\boldsymbol{x}} g_i(\bar{\boldsymbol{x}})^T \boldsymbol{d} &\leq 0 \quad \text{for} \quad i \in I(\bar{\boldsymbol{x}}), \bar{u}_i = 0, \\ \nabla_{\boldsymbol{x}} h_i(\bar{\boldsymbol{x}})^T \boldsymbol{d} &= 0 \quad \text{for} \quad i = 1, \dots, l\} \end{aligned} \tag{A.2}$$

the following holds true

$$\boldsymbol{d}^T \nabla^2_{\boldsymbol{x}\boldsymbol{x}} \mathcal{L}(\bar{\boldsymbol{x}}; \bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}) \, \boldsymbol{d} > 0.$$

Bibliography

[1] Antczak, T. (2013) Saddle point criteria and the exact minimax penalty function method in non convex programming. *Taiwanese Journal of Mathematics*, **17**(2), 559–581.

[2] Bai, X., Sun, J., Zheng, X., and Sun, X. (2012) A Penalty Decomposition Method for Probabilistically Constrained Convex Programs. *http://www.optimization-online.org/DB_FILE/2012/04/3448.pdf*.

[3] Bazaraa, M. S., Sherali, H. D., and Shetty, C. M. (2006) *Nonlinear Programming: Theory and Algorithms*. Wiley-Interscience, Hoboken, N.J., 3rd ed.

[4] Ben-Israel, A. and Mond, B. (1986) What is invexity? *The Journal of the Australian Mathematical Society. Series B. Applied Mathematics*, **28**, 1–9.

[5] Bosch, P., Jofré, A., and Schultz, R. (2007) Two-Stage Stochastic Programs with Mixed Probabilities. *SIAM Journal on Optimization*, **18**(3), 778–788.

[6] Branda, M. (2010) Solving real-life portfolio problem using stochastic programming and Monte-Carlo techniques. In Houda, M. and Friebelová, J., eds., *Proceedings of the 28th international conference on mathematical methods in economics*, pp. 67–72. University of South Bohemia, České Budějovice.

[7] Branda, M. (2012) Chance constrained problems: penalty reformulation and performance of sample approximation technique. *Kybernetika*, **48**(1), 105–122.

[8] Branda, M. (2013) On relations between chance constrained and penalty function problems under discrete distributions. *Mathematical Methods of Operations Research*, **77**(2), 265–277.

[9] Branda, M. and Dupačová, J. (2012) Approximation and contamination bounds for probabilistic programs. *Annals of Operations Research*, **193**(1), 3–19.

[10] Charnes, A. and Cooper, W. W. (1959) Chance-Constrained Programming. *Management Science*, **6**(1), 73–79.

[11] Craven, B. (1981) Invex functions and constrained local minima. *Bulletin of the Australian Mathematical Society*, **24**, 357–366.

[12] Dentcheva, D. (2009) Optimization Models with Probabilistic Constraints. In Shapiro et al. [35], chap. 4, pp. 87–153.

[13] Dupačová, J., Hurt, J., and Stěpán, J. (2002) *Stochastic Modelling in Economics and Finance*. Kluwer Academic Publishers, Dordrecht.

[14] Dupačová, J., Gaivoronski, A., Kos, Z., and Szántai, T. (1991) Stochastic programming in water management: A case study and a comparison of solution techniques. *European Journal of Operational Research*, **52**, 28–44.

[15] Ermoliev, Y., Ermolieva, T., MacDonald, G., and Norkin, V. (2000) Stochastic Optimization of Insurance Portfolios for Managing Exposure to Catastrophic Risks. *Annals of Operations Research*, **99**(1-4), 207–225.

[16] Gale, D. (1960) *The Theory of Linear Economic Models*. McGraw-Hill, New York.

[17] Han, S. P. and Mangasarian, O. (1979) Exact penalty functions in nonlinear programming. *Mathematical Programming*, **17**(1), 251–269.

[18] Hanson, M. A. (1981) On sufficiency of the Kuhn-Tucker conditions. *Journal of Mathematical Analysis and Applications*, **80**(2), 545–550.

[19] Horn, R. A. and Johnson, C. R. (1990) *Matrix Analysis*. Cambridge University Press.

[20] Janesch, S. M. H. and Santos, L. T. (1997) Exact Penalty Methods with Constrained Subproblems. *Ivestigacion Operativa*, **7**, 55–66.

[21] Kaul, R. and Kaur, S. (1985) Optimality criteria in nonlinear programming involving nonconvex functions. *Journal of Mathematical Analysis and Applications*, **105**(1), 104 – 112.

[22] Luenberger, D. (1968) Quasi-Convex Programming. *SIAM Journal on Applied Mathematics*, **16**(5), 1090–1095.

[23] Mangasarian, O. L. (1965) Pseudo-Convex Functions. *Journal of the Society for Industrial and Applied Mathematics Series A Control*, **3**(2), 281–290.

[24] Mangasarian, O. L. (1969) *Nonlinear Programming*. McGraw-Hill, New York.

[25] Mishra, S. and Giorgi, G. (2008) *Invexity and Optimization*. Nonconvex Optimization and Its Applications. Springer.

[26] Nocedal, J. and Wright, S. (2006) *Numerical Optimization*. Springer, New York, 2nd ed.

[27] Prékopa, A. (1973) Contributions to the theory of stochastic programming. *Mathematical Programming*, **4**, 202–221.

[28] Prékopa, A. (2003) Stochastic Programming. In Ruszczynski and Shapiro [32], chap. 5, pp. 267–351.

[29] Raike, W. M. (1970) Dissection Methods for Solutions in Chance Constrained Programming Problems under Discrete Distributions. *Management Science*, **16**(11), 708–715.

[30] Rockafellar, R. T. and Uryasev, S. (2002) Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance*, **26**(7), 1443–1471.

[31] Ruszczynski, A. and Frauenderfor, K. (1988) Approximation techniques in stochastic programming. In Ermoliev, Y. and Wets, R. J.-B., eds., *Numerical Techniques for Stochastic Programming*, pp. 34–64. Springer, Berlin.

[32] Ruszczynski, A. and Shapiro, A., eds. (2003) *Stochastic Programming*, *Handbooks in Operations Research and Management Science*, vol. 10. Elsevier, Amsterdam.

[33] Ruszczynski, A. and Shapiro, A. (2003) Stochastic Programming Models. [32], chap. 1, pp. 1–64.

[34] Ruszczynski, A. and Shapiro, A. (2009) Stochastic Programming Models. In Shapiro et al. [35], chap. 1, pp. 1–25.

[35] Shapiro, A., Dentcheva, D., and Ruszczynski, A., eds. (2009) *Lectures on Stochastic Programming: Modeling and Theory.* The Society for Industrial and Applied Mathematics and the Mathematical Programming Society, Philadelphia.

[36] Wolfe, P. (1961) A duality theorem for nonlinear programming. *Quarterly of applied mathematics*, **19**(3), 239–244.

List of Tables